

# Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Hyde, T. (1981) Some factors affecting journey-times on urban roads. PhD thesis, University College. [Thesis]

This version is available at: <https://eprints.mdx.ac.uk/6545/>

## Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

[eprints@mdx.ac.uk](mailto:eprints@mdx.ac.uk)

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

MX 8701896 9



THE LIBRARY  
MIDDLESEX POLYTECHNIC,  
QUEEN'S WAY, ENFIELD,  
MIDDLESEX EN3 4SF

UNIVERSITY COLLEGE

TRANSPORT STUDIES GROUP

LONDON

SOME FACTORS AFFECTING JOURNEY-TIMES ON URBAN ROADS

A thesis submitted for the award of the Degree of Doctor of Philosophy,  
University of London by

T. Hyde, BSc, CEng, MICE, MIMunE, MIHE, ACIT.

April 1981

THE LIBRARY  
MIDDLESEX POLYTECHNIC,  
QUEEN'S WAY, ENFIELD,  
MIDDLESEX EN3 4SF

SITE <b>B8</b> HE	MIDDLESEX POLYTECHNIC LIBRARY
ACCESSION No.	8701896
CLASS No.	388.4131
SPECIAL COLLECTION	THESIS COLLECTION

H4D

05991834

#### ACKNOWLEDGEMENTS.

I gratefully acknowledge the penetrating comments and helpful advice given by my tutor Mr. J.G.Wardrop, B.A.,F.S.S., Reader in Traffic Studies, University College, London.

I am particularly indebted to Mesars. A.Wilderspin and K.Johnson, laboratory technicians at Middlesex Polytechnic, for their considerable assistance with the fieldwork and also to my wife, Helen, for her assistance with fieldwork and analysis of cine-films. Finally, I thank my employers, Middlesex Polytechnic, who sponsored this research and provided laboratory facilities.

# ABSTRACT.

This thesis discusses the results of two types of study, (a) of the relationships between journey-time and traffic flow during peak periods on two sections of urban road, each about 0.4 miles in length, and (b) of relationships between journey-times and a number of land-use variables over about 50 miles of suburban main roads in off-peak periods.

The results of study (a) show that the most significant relationships occurred when the data was analysed using non-linear or two-regime models. In the latter case the two regimes were separated by a critical flow. Above the critical flow changes in flow were correlated with changes in journey-time and below the critical flow the changes appeared to be random. Further evidence is provided to show that the use of different sampling intervals can give rise to different empirical relationships with the same data. A queueing model postulated by Davidson (1968) has been shown to give a satisfactory agreement with the results of study (a). A model is developed in this thesis which accounts for consistent changes in saturation flow arising from the use of different sampling intervals in Davidson's model. Additionally, an empirical index has been derived which reflects the changes in activity during the peak hour, and journey-time is highly correlated with this index.

Study (b) shows that after the effects of major intersections have been accounted for, journey-time was significantly correlated with a number of factors of land-use along the suburban main roads which were studied. Furthermore, a number of mathematical techniques have been used which permit the calculation of indices of the variation of the

surrounding conditions and these techniques were shown to produce robust and repeatable results. Journey-time was significantly correlated with an Activity Index which was associated with a number of land-use parameters.

## CONTENTS

### ACKNOWLEDGEMENTS

### ABSTRACT

### List of Tables

### List of Figures

Chapter 1	Introduction	Page 15
Chapter 2	Historical Review	Page 19
Chapter 3	Peak Hour Studies	Page 32
3.1	Survey Sites and Methods	Page 32
3.2	Results and Analysis	Page 35
Chapter 4	Two-Regime and Non-Linear Models	Page 50
4.1	Discontinuous Models	Page 50
4.2	Two-Regime Models	Page 54
4.3	Davidson's Model	Page 55
Chapter 5	Multivariate Analysis of Speed-Flow Data	Page 65
5.1	Application to Speed-Flow Data	Page 68
5.2	Development of Factor Models	Page 72
Chapter 6	Off-Peak Journey-Time Studies	Page 82
6.1	Sites, Methods and Preliminary Analysis	Page 82
6.2	Analysis	Page 87
Chapter 7	Multivariate Analysis of Off-Peak Data (cont.)	Page 100



## CONTENTS

7.1	Factor Analysis Models	Page 100
7.2	Regression Analysis of Factor Scores	Page 107
Chapter 8	Summary and Conclusions	Page 123
8.1	Journey-Time/Flow Studies	Page 123
8.2	Off-Peak Journey-Time Studies	Page 127
8.3	Suggestions for Further Work	Page 130
Appendix A	References	Page 132
Appendix B	Tables	Page 139
Appendix C	Figurea	Page 177
Appendix D	Computer Programs	Page 201
Appendix E	Principles of Factor Analysis	Page 206

# C O R R I G E N D A

- Page 38 Replace the second sentence by the following:  
 "They were only partially observed in the data collected in the off-peak period on 17/7/72 where the slope regression coefficients were much more varied. But the coefficients were not significantly different from zero in most cases."
- Page 52 Add to the end of the first sentence:  
 "for this data."
- Page 79 Add to the penultimate sentence:  
 "the result is highly significant for the westbound flow and non-significant for the eastbound flow."
- Page 39 Change "Table 3.2a" to "Table 3.3"
- Page 41 Change "Fig. 3.8" to "Fig. 3.9"
- Page 75 Change " $Y$  = mean value of  $Y$ " to " $\bar{Y}$  = mean value of  $Y$ "
- Page 148 In Table 3.8 the following labels should be interchanged:  
 "Flow and Lagged Journey-Time" , "Flow and Lagged Flow"
- Page 167 Table 6.14 should be replaced by a new copy of Table 6.14 shown below.

Table 6.14

Linear Regression Models: Off-Peak Journey-Time Study

- Total Data (N = 99)

Dependent Variable	Model	Correlation Coefficient	Remarks
Journey-Time (Y) (min)	$Y = 1.09 + .159 X_S$	.619	Highly Significant
	$Y = 1.06 + .106 X_S + .0037 X_5$	.694	-ditto-
	$Y = 0.95 + .103 X_S + .0035 X_5 + .027 X_B$	.716	-ditto-
	$Y = 1.26 + .127 X_S - .0025 X_4$	.667	-ditto-

## LIST OF TABLES

All tables are located in Appendix B.

Table 3.1	Page 140
Survey details, Journey-time/Flow Study.	
Table 3.2	Page 142
Linear Regression Analyses of High Street Data.	
Table 3.3	Page 143
Slope Regression and Correlation Coefficients of the Relationships between Flow and Flow-Weighted/Unweighted Journey-Times for Southbury Road, (Westbound) 19/10/73.	
Table 3.4	Page 144
Effects of Changing Sampling Intervals.	
Table 3.5	Page 145
Linear Regression Analyses of Southbury Road Data.	
Table 3.6	Page 146
Correlation Matrix of Variables Used in Multiple Linear Regression Analysis - High Street, 19/7/72, (5 Minute Intervals).	
Table 3.7	Page 147
Multiple Linear Regression Results - High Street, 19/7/72	
Table 3.8	Page 148
Lagged Serial and Cross-Correlations between Flow and Journey-Time, High Street, 19/7/72.	
Table 3.9	Page 149
Slope Regression and Correlation Coefficients Predicted by Wright's Model Compared with Values Calculated from the Aggregated Data, (Westbound), 19/10/73.	

## LIST OF TABLES

Table 3.10	Maximum Values of Slope Regression and Correlation Coefficients Predicted by Wright's, (All Peak Period Data).	Page 150
Table 3.11	Maximum Slope and Correlation Coefficients Predicted by Wright's Models, High Street, 19/7/72.	Page 151
Table 3.12	Relationships between Journey-Times and Pseudo-Concentration, 5 Minute Averages.	Page 152
Table 4.1	Combined Sum of Squares of Discontinuous Model Between Flow and Journey-Time (5 Minute Averages), High Street, Total Data.	Page 153
Table 4.2	Combined Sum of Squares and Residual Squares of Relationships between Mean Flow and Journey-Time, High Street, Total Data.	Page 153
Table 4.3	Combined Sum of Squares and Residual Squares of Relationships between Mean Flow and Journey-Speed, High Street, Total Data.	Page 154
Table 4.4	Discontinuous Regresaion using a Power Curve on High Street Combined Data - Relationship between Journey-Time and Flow.	Page 154
Table 4.5	Analysis of High Street Total Data using Dsvidson's Travel Time Model to Predict Journey-Times.	Page 155
✓ Table 4.6	Saturation Flows - High Street using Davidson's, Poisson and Normal Distributions.	Psge 155
Table 5.1	Correlation Matrix of Data Collected on 17/7/72 in the High Street, Off-Peak.	Page 156

## LIST OF TABLES

Table 5.2	Page 156
Correlation Matrix of Data Collected on 19/7/72 in the High Street, Peak.	
Table 5.3	Page 157
Eigenvalues and Percentage of Variation Explained by the Principal Components of Variables X1, X2, X3 and X4 - High Street.	
Table 5.4	Page 158
Loadings on Two Principal Components in the Peak Period, High Street, 19/7/72. 5 Minute Data, N = 18.	
Table 5.5	Page 158
Effects of Data Aggregation into Successively Longer Sampling Intervals on the Eigenvalues of the Principal Components, High Street, 19/7/72.	
Table 6.1	Page 159
Parameters Measured in Journey-Time Study, Enfield (1979), (Single-Carriageways Only) - 45 Sections.	
Table 6.2	Page 160
Parameters Measured in Journey-Time Study, Haringey (1979), (Single-Carriageways Only) - 15 Sections.	
Table 6.3	Page 161
Parameters Measured in Journey-Time Study, Barnet (1980), (Single-Carriageways Only) - 39 Sections.	
Table 6.4	Page 162
Results of Linear Regression Analysis of Data Collected on Dual-Carriageways in Enfield - 1978 (Sample Size = 15).	
Table 6.5	Page 162
Regression Models: Single Carriageways - Enfield Data. (N = 45).	
Table 6.6	Page 163
Regression Models: Single Carriageway Data (N = 42), Enfield 1979.	

## LIST OF TABLES

Table 6.7	Correlation Between Observed and Predicted Journey-Times and Running-Speeds, Using Models Derived from the 1978 Data on the 1979 Data Collected in Enfield.	Page 163
Table 6.8	Regression Models: Single-Carriageway Data, Haringey 1979, (N = 15).	Page 164
Table 6.9	Comparison of Models Produced by Enfield 1978, 1979 and Haringey 1979 Data.	Page 164
Table 6.10	Correlation Between Observed and Predicted Journey-Times and Running-Speeds, Haringey 1979, Using Models Derived from Enfield (1978) Data.	Page 165
Table 6.11	Regression Models: Single Carriageway Data, Barnet 1980 (N = 39).	Page 165
Table 6.12	Correlation Matrix - Barnet Off-Peak Journey-Time Study 1980.	Page 166
Table 6.13	Linear Regression Models, Barnet (1980) Using Variables XS, X4 and XB only.	Page 167
Table 6.14	Correlation Between Observed and Predicted Journey-Times and Running-Speeds, Barnet 1980 Using Models Derived from Enfield (1978) Data.	Page 167
Table 6.15	Regression Models of Spatially Aggregated Journey-Time Data. Showing Effects of Aggregation on the Coefficients. Total Data.	Page 168
Table 7.1	Factor Loadings Obtained by Principal Component Analysis and Factor Analysis of the Pooled Off-Peak Journey-Time Data.	Page 169

## LIST OF TABLES

Table 7.2	Page 170
Correlation Matrix of Rotated Factors and Journey-Times for the Pooled Off-Peak Data.	
Table 7.3	Page 171
Most Significant Linear Regression Equations between Factor Scores and Off-Peak Journey-Times/Journey-Speeds. Pooled Data.	
Table 7.4	Page 172
Variables Retained by Variable Elimination Processes Types B1 and B2, and Results of Multiple Linear Regression Analysis of the Remaining Variables.	
Table 7.5	Page 173
Multiple Correlation Coefficients of Each Variable with Remaining Variables, Variable Elimination Process Type A2 - Off-Peak Journey-Time Study, Total Data.	
Table 7.6	Page 174
Communalities of Nine Variables Based on the Factor Loadings for Two Rotated Factors: Pooled Data.	
Table 7.7	Page 175
Variation of Factor Loadings of First Two Factors as the Number of Variables is Reduced: Pooled Data.	
Table 7.8	Page 176
Details of Multiple Regression Models of Normalised Journey-Time and Factor Scores of Reduced Variable Models: Pooled Data.	

## LIST OF FIGURES

The figures are to be found in Appendix C.

Fig. 2.1	Relation Between Mean Speed and Flow in Central London.	Page 178
Fig. 2.2	Form of Empirical Travel Time-Flow Relationships for Observation Intervals of Various Durations - Rothrock (1957).	Page 179
Map 3.1	Section of High Street, Ponders End Investigated in the Peak-Period Studies.	Page 180
Map 3.2	Section of Southbury Road, Enfield Investigated in the Peak-Period Studies.	Page 181
Fig. 3.1	Variations of Flows and Journey-Times, High Street (Southbound), 19/7/72 Evening Peak Period.	Page 182
Fig. 3.2	Variations of Flows and Journey-Times, Southbury Road (Eastbound), 19/10/73 Evening Peak Period.	Page 183
Fig. 3.3	Variations of Flows and Journey-Times, Southbury Road (Eastbound), 5/12/73 Morning Peak Period.	Page 184
Fig. 3.4	Variations of Flows and Journey-Times, Southbury Road (Westbound), 5/12/73 Morning Peak Period.	Page 185
Fig. 3.5	Scatter Diagram of Mean Journey-Times and Mean Flows (5 Minute Averages), High Street - 19/7/72.	Page 186



## LIST OF FIGURES

- |          |   |          |
|----------|---|----------|
| Fig. 3.6 | Scatter Diagram of Mean Journey-Times and Mean Flows<br>(5 Minute Averages), Southbury Road (Westbound) -<br>5/12/73  | Page 187 |
| Fig. 3.7 | Scatter Diagram of Mean Journey-Times and Mean Flows<br>(5 Minute Averages), Southbury Road (Eastbound) -<br>19/10/73.  | Page 188 |
| Fig. 3.8 | Scatter Diagram Of Mean Journey-Times and Mean Flows<br>(5 Minute Averages), Total Data - High Street.  | Page 189 |
| Fig. 3.9 | Scatter Diagram of Standardised Journey-Time Residuals<br>and Observed Journey-Times, Total Data - High Street.   | Page 190 |
| Fig. 4.1 | The Discontinuous Linear Regreasion Relationship.   | Page 191 |
| Fig. 4.2 | Effects of Changes in J - Davidson's Function.  | Page 192 |
| Fig. 4.3 | Three Alternative Models of Journey-Time/Flow<br>Relationships.   | Page 193 |
| Fig. 5.1 | Plot of Four Variables on Axes of the Two Most<br>Significant Principal Components Showing the Axes<br>after Varimax Rotation - High Street, 5 Minute Data,<br>19/7/72. | Page 194 |
| Fig. 5.2 | Scatter Diagram of Predicted and Observed<br>Journey-Times Using Factor Model.  | Page 195 |
| Fig. 6.1 | Scatter Diagram of Observed and Predicted<br>Journey-Times Enfield 1979 based on 1978 Model.  | Page 196 |

## LIST OF FIGURES

- Fig. 6.2 Page 197  
Scatter Diagram of Observed and Predicted  
Journey-Times in Haringey Based on the Enfield 1978  
Model.
- Fig. 6.3 Page 198  
Scatter Diagram of Off-Peak Journey-Times - Total  
Data, Linear Regression Model.
- Fig. 7.1 Page 199  
Vector Diagram of Factor Loadings of Nine Variables on  
Two Factors, Showing Positions Before and After  
Orthogonal Rotation - Off-Peak Journey-time Study,  
Total Data.
- Fig. 7.2 Page 200  
Scatter Diagram of Observed and Predicted  
Journey-Times, Off-Peak Journey-Time Study, Total  
Data, Factor Model.

## CHAPTER 1

### INTRODUCTION

This research is concerned with an investigation into the effects on journey-times and journey-speeds on urban roads of a number of factors which are expected to influence them, such as traffic flowa, pedestrian flows, parked vehicles and adjacent land-uses. The choice between journey-speeds or journey-times as parameters for investigation is discussed by the Bureau of Public Roads (1965). Journey-time is the concern of the driver when planning a journey, rather than journey-speed, and traffic engineers must be able to determine journey-times when assessing the benefits of highway proposals or when carrying out traffic assignment in transportation studies. Furthermore journey-speeds are not additive over several sections of the same journey whereas journey-times are. However, there are certain advantages in studying journey-speeds since, under free-flow conditions, the distributions tend to be symmetrical whereas journey-time distributions exhibit a long positive "tail". Thus assumptions of the normality of distributions are more likely to be valid for journey-speeds than journey-times and the accuracy of the predictions of journey-speeds can be tested.

## INTRODUCTION

The fieldwork was carried out on a number of main roads in the London Boroughs of Enfield, Haringey and Barnet and the details will be given in a later chapter. Insofar as these roads are typical of roads in urban areas then the results and conclusions may be expected to apply elsewhere and where unique or unusual effects occur these will be identified.

The studies have been carried out in both peak and non-peak periods, the major difference between them being that in peak periods variations of flow and journey-time may be significantly correlated whereas in non-peak periods the variations are independent.

In the peak period studies two sections of roads in Enfield have been investigated in detail. Data concerning journey-times, traffic flows, turning vehicles, crossing pedestrians and traffic signal times has been collected. The basic sampling period was a one-minute interval and some of the important conclusions of the study are concerned with the effects of different sizes of sampling period on the results when the data is aggregated over successively larger sampling periods.

In off-peak periods journey-times were found to be independent of traffic flows and consequently different types of studies were carried out on these and other roads at these times. It is clear that if there are differences in journey-speeds on different roads in off-peak conditions then these differences must arise from variations in the characteristics of the features along the road such as the frequency of intersections, pedestrian crossings, parked vehicles, adjacent land-uses and, possibly, weather. Extensive studies have been carried out to observe and measure these effects and the results will be discussed in

## INTRODUCTION

later chapters. Often it is not evident which variables are likely to be important and sometimes the variables measured may be acting as proxy for the factors which may be most significant in influencing journey-time. This possibility was considered when the studies were planned and the methods of analysis used were those which could take account of these effects. The study relies upon established methods of multiple regression analysis in order to detect any predictive relationships. In addition other methods of multivariate analysis have been used in order to deal with some of the problems caused by intercorrelations between the "independent" variables. Attention will be directed to the use of transformations of data used in other studies which may be invalid because of intercorrelations which these transformations create. Problems caused by the lack of standardisation of sampling periods in different studies - for example, certain kinds of bias and difficulties in making valid comparisons - will also be discussed.

It is implicit in this kind of research that great use must be made of computing facilities and some of the analyses carried out were only made possible because of the new computing system which was installed at Middlesex Polytechnic in 1977. The program packages used and the special programs written for this study are briefly described in Appendix D to this thesis.

In order to place this study in perspective the next chapter reviews the historical development of a number of theories about the factors affecting journey-times and journey-speeds and the various empirical studies which have been carried out. Such an historical review is also

## INTRODUCTION

useful in identifying gaps in the knowledge of traffic behaviour and to comment on the methods and results of different studies.

## CHAPTER 2

### HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

Relationships between the fundamental parameters of traffic flow have been of considerable interest to traffic engineers for many years. Both theoretical and empirical relationships have been postulated with varying levels of reliability in the resulting predictions. Smeed (1949) reported the results of investigations into the relationships between both running and journey-times of traffic and flows on roads in Central London. On a number of roads of average width 48ft a linear relationship was found between running speed and vehicle flow with a maximum speed of 26.8 mph. The slope of the line was 0.73 mph/100 vph and this relationship was found to apply over a range of flows from 370 to 1740 vph. In the same report the maximum journey-speed was reported to be 25.3 mph and the regression slope was 0.81 mph/100 vph; the lengths of the routes investigated varied between 0.13 and 0.75 miles. The periods of the time sampling intervals (for flow measurements) were not stated and it is likely that these varied between sections since the moving-observer method was used for collecting the data. A number of later studies, Rothrock and Keefer (1957), Wright (1972), have shown that the use of different sampling periods can produce differences in the regression coefficients and it is advisable to report the sampling intervals in any such study.

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

Further results of investigations in Central London were reported by Newby et al (1950) and Wardrop (1952) which applied to roads over a wide range of widths and traffic flows. The results were expressed by means of the following equation:

$$V = 31 - \frac{Q + 430}{3(w - 6)} \quad ; \text{ or } 24 \text{ mph}$$

whichever is less,

where  $V$  = running speed (mph)

$Q$  = traffic flow (vph)

$w$  = road width (feet)

The results were also represented graphically which shows more clearly the nature of the relationship for typical road widths (see Fig 2.1). In R.R.L. (1965) these results are shown as the basis for capacity-restrained traffic assignment and the equation will be further discussed later in this chapter.

Rothrock and Keefer (1957) report the results of an investigation into the relationships between journey-time and entering traffic flow in a congested street in Charleston; the investigation was concerned with the measurement of urban congestion using an empirical index of congestion calculated from total vehicle occupation time. The basic data was collected at both ends of the street and certain intermediate points along the section over a series of 6 minute sampling intervals. No reason was given for the choice of entering flow as abscissa although this could have been important since the length of the street investigated included two intermediate cross-roads and a signal controlled cross-road at the upstream end of the street. Edie (1963)



## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

argues that the flow of a traffic stream should ideally be defined as the aggregate distance travelled by all vehicles passing through a space-time domain divided by the area of the domain. However, this is a difficult measure of traffic flow to observe accurately and few studies since 1963 (and none before) have reported using this measure of flow. An interesting feature of Rothrock's results was the observation of the turning back effect (see Fig. 2.2) on the journey-time/flow curves at certain times. This effect was postulated by Greenshields (1952) and other authors - for example, see Wardrop (1963) - although usually the argument for the existence of such an effect is made from the point of view of traffic flow and spot speed relationships. However, as will be argued later, this effect may only be observable where congestion is caused at some intermediate point and entering flow is affected, - see Branston (1976). One other important feature of Rothrock's results related to the present study is the use of alternative time intervals for displaying the results. These were shown in graphs corresponding to time intervals of 6, 12, 18, 24, 30 and 60 minutes respectively and are reproduced in Fig 2.2. The effects of different aggregations on average values of the parameters were not discussed in that paper, although this kind of aggregation can have a significant effect on the relationships obtained.

An interesting study was carried out by Guérin (1958) into the relationship between traffic flow and journey-time in Washington Boulevard, Chicago and Temple Street, New Haven. Elevated observers measured vehicle journey-times and entering and leaving traffic flows in the test sections. The data were collected in 1 minute sampling intervals and aggregated in a number of different ways depending upon

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

the variable being aggregated and the traffic conditions during the period being considered. When the travel time was less than 2 minutes the mean traffic flow was defined as the average of the entering flow during the previous 2 minutes and the leaving flow during the following 2 minutes. When the travel time was greater than 3 minutes the flow was defined as the weighted average of the entering flow during the previous 3 minutes and the leaving flow during the following 3 minutes. The flows in the 3 consecutive minute periods were weighted 1, 2, 1. Concentrations were aggregated in a similar way. The concentrations were measured indirectly by assuming that concentration is equal to flow times average travel time. (This will later be called pseudo-concentration in this report). High correlations were reported between journey-time and concentration (0.94 and 0.93 respectively). The correlations between journey-time and flow were insignificant in both studies.

Underwood (1960) postulates an exponential relationship between space-mean speed and concentration. Data from two sites were used to justify this hypothesis. As a result of these investigations it was suggested that a generalised speed-volume diagram could be used to identify three separate zones of behaviour; a zone of normal flow, a zone of unstable flow and a zone of forced flow. No information was provided on the practical methods or assumptions used in this study and, as will be discussed later, these can have a significant effect on the conclusions.

An unusual study was reported by Wardrop (1963) of experimental speed/flow relations in a single lane. The experiments were carried out on circular tracks of radii approximately 50 ft, 100 ft, 200 ft, and 400 ft respectively and on a straight track. In some of the

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

experiments the number of vehicles was fixed and they were driven as fast as was comfortable, whilst on the straight the speed of one vehicle was fixed and the others followed at minimum spacing. The relationships between speed and flow indicated a certain degree of agreement with those found on actual roads where speeds were comparatively high and using the headway relationship quoted by Smeed and Bennett (1949). The relationship observed between speed and concentration was similar to that postulated by Underwood (1960).

A flow/travel time relationship derived from queueing theory was postulated by Davidson (1966),

$$T = T_0 + T_0 \frac{CJ}{1-C}$$

where  $T$  = journey-time

$T_0$  = journey-time at zero flow

$C$  = ratio of arrival rate/service rate,

which is equivalent to the ratio

of arrival flow/saturation flow

$J$  = delay factor (sometimes called Davidson's factor)

Davidson suggested the use of the factor  $J$  to account for some of the differences between the multiple queueing situation which occurs on a finite length of road and the idealised single-server queueing system from which the model was derived. The value of  $J$  can be varied depending upon the type of road and the associated factors which give rise to delay on that road. The model was tested against results published by Irwin et al (1962) and Wardrop (1952). It is not possible to use linear regression to determine the significant constants in the relationship and a method of successive

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

approximations was adopted in order to minimise the mean square error. When plotted, the postulated curves were seen to fit the observations satisfactorily for both sets of data. The analysis of Wardrop's data was more revealing because of the wider range of types of roads and traffic flows which were observed. The parameter  $J$  as calculated by this process, varied inversely with road width. In cases where high values of  $J$  apply even at relatively low flows significant increases in journey-time occur for increasing service rates. Thus the model can describe observed conditions quite closely.

One of the criticisms levelled at Davidson's model is that it predicts that very large journey-times will occur when the traffic flows approach the saturation flow, whereas in reality the flows may be small when the journey-times become very large. Another difficulty is that traditional queueing theory does not account for the space occupied by the queue. The queue could be assumed to occur at the downstream end of the link and the traffic flow measured at the upstream end, but in short links the queue can extend for the full length of the link. Suitable values of the model parameters are difficult to determine, the method of successive approximations has been used by Davidson and the use of the model has only been reported in Australia. Taylor (1977) reports a new method of estimating  $J$  using the least-squares principle which may lead to greater use of this model.

Speed/flow relationships in an urban area were discussed by Dick (1968) in a paper which reported the results of such studies in Newcastle. The important contribution of this paper was that it extended the studies reported by Wardrop (1952) to take account of differences in the surroundings through which the roads passed,

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

although the results were limited because the roads were categorised into only four groups. The groups were determined by considering two major factors - number of parked vehicles/mile and number of major intersections /mile. Journey-times were measured by driving a vehicle round a route keeping pace with the general traffic flow. Traffic flows were measured using a traffic counter in each section. No details were given of the sampling intervals or the lengths of the sections of road which were investigated. The results were shown in several speed/density scatter diagrams using density (concentration) derived by dividing the volume (flow) by the speed. High correlation coefficients were reported for all of the roads but this is invariably the case if concentration is derived in this way and then used as an "independent" variable. No scatter diagrams of speed and flow were shown although idealised relationships were displayed in a graph. The results indicated that significant differences occurred between the journey speeds on different types of road at times of low flows.

In a paper concerned with journey-speed and flow in central urban areas Wardrop (1968) derived a semi-empirical formula relating journey-speed to a number of characteristics of roads such as road width, traffic flow, number of controlled intersections per mile of road and the proportion of green-time. He argued that the journey-time per mile was composed of two parts, the running time per mile and the delays caused by controlled intersections. Thus:

$$\frac{1}{V} = \frac{1}{V_r} + FD$$

where  $V$  = journey-speed (mph)

$V_r$  = running-speed (mph)

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

D = average delay per intersection (hours)

F = number of intersections per mile.

The values of running-speed were based on the results of investigations previously reported by the Road Research Laboratory (1965),

$$V_r = 31 - \frac{0.70 + 430}{3w}$$

where Q = flow (pcu/hour)

W = carriageway width (feet)

It was shown that, for both vehicle actuated and fixed time signals, the reciprocal of delay at signal controlled intersections varied approximately linearly with degree of saturation over a wide range of flows. (Degree of saturation is the ratio of flow to saturation flow). This is equivalent to;

$$D = \frac{B}{1 - Q/LS}$$

where B = constant

LS = absolute capacity of intersection.

The resulting equation for journey-speed is;

$$\frac{1}{V} = \frac{1}{31 - \frac{140}{W} - \frac{.024Q}{W}} + \frac{F}{1000 - \frac{6.8Q}{LW}}$$

The formula assumes that the differences between opposing flows on two-way roads have no significant effect on journey-speed; this may be true at low flows but on roads where tidal flows occur the errors may be significant at peak periods because of the large increase in

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

delays at signal controlled intersections. This equation is similar to that of Davidson since both equations are composed of two terms, one for journey-time at low flows and the other for delays at intersections. Furthermore, the divisor in the delay term is identical in both equations.

The relationship between journey-speed and flow has been investigated at several sites on main roads in suburban areas by Freeman Fox and Associates (1972). The investigations were carried out on principal roads and trunk roads over sections of about two miles; the moving-observer method was used for measuring journey-times and flows were determined by simultaneous roadside counts at a number of intermediate points. The data was collected on weekdays only and included details of adjacent land-uses, parked vehicles, degree of access, pedestrians, rise and fall, and curvature, in addition to speeds and flows. The data was classified for different road types and vehicle types. The results of the investigations were not very satisfactory, a number of multiple regression equations being obtained several of which had significant regression coefficients of the "wrong" sign, for example, on one road the equations predicted increases in journey-speed with increases in goods vehicle flow. The weather condition variable (measured on a subjective severity scale) gave reasonable and consistent results. Apart from this variable and the flow of light vehicles, the regression coefficients were stated to be unreliable, possibly because of the inter-correlations between the explanatory variables. It was intended to obtain a separate regression equation for each class of vehicle but this was abandoned and a result for a standard traffic stream composition was obtained. The relations obtained are as follows;

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

$$V = V_0 + \frac{S(Q - 300)}{1000}$$

where  $V$  = journey-speed (kph)

$Q$  = one-way vehicle flow (vph), per standard lane.

$V_0$  is the "free speed", which is the speed of the stream when  $Q$  is 300 vph/lane. When it is impossible to measure  $V_0$  directly the report suggests that the following equation should be used;

$$V_0 = (5 + \frac{D}{10}) - 10(I - .8) - .167(A - 27.5)$$

where  $D$  = proportion of dual carriageway

$I$  = number of major intersections/km

$A$  = accessibility (number of accesses/km)

$S$ , the slope of the equation, is given by:

$$S = -25 - 1.333(V_0 - (50 + \frac{D}{10})) - 30(I - .8) - .4(B - 65)$$

where  $B$  = proportion of roadside that is developed.

These equations explained 55% and 62% respectively of the variance of the dependent variables.

High variability of the basic data collected in traffic studies can cause the problem of ill-defined relationships, this is a particular difficulty in studies of dual-carriageways. Duncan (1974) discusses the results of such a study in which speeds were apparently not affected by flows except at very high levels.

In recent years the most comprehensive paper to be written on this area of study was that of Branston (1976). This paper was concerned with the measurement and formulation of link capacity functions, or volume-delay curves, for use in traffic assignment procedures. In the



## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

first part of the paper data collection methods were discussed and, in particular, the inconsistencies which can arise from the choice of the position at which flows are measured and the length of the sampling intervals used. He argued that if the downstream end of the link had a lower capacity than the upstream end, and the entry flow exceeded the exit capacity, then the link journey-time would increase as vehicles were forced to queue in the link. The maximum journey-time would correspond to the situation where the queue extended over the whole link and thus the entry flow would then be reduced to the downstream capacity. The flow/journey-time relationship would show the turning-back effect if the entry flow was used to derive the relationship but not if the exit flow was used. He concluded that the turning-back effect observed by Rothrock and Keefer (1957) would only be pronounced for short sampling intervals and if the data were aggregated over longer intervals the effect would become gradually less pronounced. He also stated that journey-speed flow relationships should be measured under steady-state conditions, when entry and exit flows are identical. In general, if the entry flow remains constant for some period this can be regarded as a steady-state condition. For a long link, where the exit queue does not reach the entry point, flows should be measured at the exit point in order to assess the effect of capacity constraint on flow. The latter two relationships may be difficult to apply to capacity-restrained assignment if the assigned flow exceeds capacity because the curves approach infinite journey-times at capacity. The curves postulated by the Bureau of Public Roads (1964) and Steenbrink (1974) were also discussed. These curves were criticised on the grounds that they require individual calibration for each application. There was a need for empirical relationships whose parameters could be defined by characteristics

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

such as speed limits, street widths and traffic signal settings in order to avoid extensive data collection. There appeared to be only one example of such a model, namely that derived by Wardrop (1968), this had a theoretical basis although part of the model was based on empirical observations of traffic behaviour.

A paper by Bampfylde et al (1978) describes the analysis of data concerning speeds and flows of traffic in tunnels. The authors identify the major difficulties of producing satisfactory relationships since there are few examples of tunnels of significant lengths and because traffic is sometimes constrained to travel in specified lanes without overtaking. Thus comparisons with other roads may be invalid, although some comparisons were made with the results quoted by Duncan (1974) and Freeman, Fox and Associates (1972). The most obvious feature of Bampfylde's results was the wide range of scatter of the data for individual tunnels and the wide range of regression coefficients between tunnels. The sampling periods were given as 5, 10 and 15 minutes for the measurement of flows and, although it is likely that these were maintained constant for all of the studies, it is not clear which sampling intervals were used to derive the empirical relationships. Neither was it stated whether different sampling intervals led to identical relationships for each tunnel. Although the differences in lengths of the tunnels were noted, between 628 and 3138 metres, the effects which may be related to differences in length were thought to be end effects or the reaction of drivers to the proximity of side walls. The possibility of spatial cross-correlations in the data, comparable to the serial correlations discussed by Wright (1971), was not considered. The authors concluded that the considerable differences in the results did not permit a

## HISTORICAL REVIEW OF FLOW/JOURNEY-TIME STUDIES

separate relationship to be established for traffic flow in tunnels.

Duncan (1979) points out other dangers of mis-placed confidence in implied speed/flow relationships derived from apparently well-defined speed/concentration relationships. A set of spot-speed and mean concentration data which had been collected over a period of 5 minutes was analysed. This showed that estimates of saturation flow between 2230 and 3060 vehicles/hour could be obtained depending upon the form of the assumed relationship between concentration and speed. However, much of this variation could be attributed to an obvious discontinuity in the speed/concentration data and this might have been an unusual feature even for this particular location since the period of observation was short. This paper is a valuable addition to the growing evidence of the inacceptability of transforming data from the space domain to the time domain and vice-versa - see also Weiner (1974) and, in particular, Braiman (1972). The consequences of this transformation will be discussed later in this thesis using data collected for this study.

The investigations reported in this thesis commenced in 1972 and the earliest investigations were centred on the type of problem which was important at that time - the relationships between journey-time and traffic flow. Later studies were concerned with the relationships between journey-times and a number of other parameters when it was observed that variations of journey-time and variations of flow were not correlated in off-peak periods. Some of the most recent papers discussed in this chapter have had some influence on the nature and direction of the present studies and these papers will be considered where relevant. In the next chapter the earliest of the author's investigations will be described.

## CHAPTER 3

### FLOW/JOURNEY-TIME STUDIES

These studies were concerned with the detailed investigation of the relationships between journey-times and traffic flows on two sections of roads in Enfield. Similar studies were carried out on these roads in peak periods and between peak periods using identical methods and these will be discussed in this chapter. The periods between the peaks will be called off-peak periods in this thesis. A later chapter will discuss other types of off-peak studies concerned with the variations of journey-time between different roads.

#### 3.1 SURVEY SITES AND METHODS.

##### Sites

The first site is the section of the High Street, Ponders End, Enfield between Southbury Road and Lincoln Road (see Map 3.1). The section is 530 yards long and each end is subject to traffic signal control. The road has a moderate "S" bend which does not permit a continuous view between the ends. There are two intermediate intersections and a zebra crossing along the section. Parking is not permitted, but a number of commercial vehicles stop to load and unload goods at the numerous shops and other premises. There are two bus-stops serving each direction. The average carriageway width is about 27 feet and varies between 24 and 36 feet. The mean off-peak journey-time of

## FLOW/JOURNEY-TIME STUDIES

traffic on this road was found to be about 1.24 minutes and the corresponding mean journey-speed was 14.6 mph.

The second site is a section of Southbury Road from a point about 100 yards west of the A10 intersection to a bus-stop 20 yards west of Ladysmith Road, a distance of 880 yards (see Map 3.2). This section has a moderate bend which prevents a continuous view between the ends and there are six intermediate T-junctions on each side of the road. There is one intermediate bus-stop serving each direction. The average carriageway width is about 30 feet and is practically constant throughout the whole length. Both sides are mostly lined with houses but there are a few shops at the western end of the section, a petrol station on the south side and a school on the north side. There is a zebra crossing near the mid-point of the section. The mean off-peak journey-time is 1.05 minutes and the corresponding mean journey-speed is 28.6 mph.

### Methods

Table 3.1 gives the details of the surveys carried out at each of these sites. The first two surveys used a team of eight observers in order to collect data. Journey-times were measured using the registration number method at a theoretical sampling rate of 50%. Classified traffic counts, green times and cycle times, numbers of pedestrians using the pedestrian crossing and numbers of vehicles turning at the intermediate T-junctions were also recorded.

The data were collected and recorded in one-minute sampling intervals. Each observer used a synchronised stop-watch for recording times. The stop-watches were checked for inaccuracies after each survey. Vehicle times, measured to 1/100th of a minute, and registration numbers were

## FLOW/JOURNEY-TIME STUDIES

recorded on a portable tape recorder. After preliminary analysis of the first two surveys it was decided that the registration number method would not be suitable for measuring the journey-times at higher flow-rates because of the relatively low sampling proportions which were obtained. The proportions of vehicles whose journey-times were measured in these two surveys were 17% and 24% and in some intervals no measurements of journey-time were obtained. This could have been because either the vehicles which passed did not have the required digits, or occasionally because of human errors or because no vehicle passed both observation points during these intervals.

Later surveys used the cine-photography technique for recording much of the data at the ends of both sites. It was found to be impracticable to synchronise the film speeds exactly at the two points but the need for this was avoided by simultaneously recording the frame numbers and stopwatch times of a number of easily identified events. A manual record of these times and a description of the events were made and used later for calibration of the film play-back in the laboratory. It was found too difficult to observe a stopwatch on film, so a continuous and simultaneous film record of times was not made. If this had been possible the analysis of the films would have been simplified. The filming rate which was used was 50 frames per minute and this gave about 80 one-minute samples of data from the longest available film. After developing, the identifiable events were located on the films and their recorded frame numbers were used to determine a datum frame on each film. The analysis of the film was carried out by setting the frame counter to zero on the projector when the datum frame was projected.

## FLOW/JOURNEY-TIME STUDIES

The analysis of each pair of films was usually carried out simultaneously in order to identify each vehicle. The corresponding frame numbers were recorded at the appearance of a vehicle in frame and converted later to elapsed time. The average time taken to analyse a pair of films, each of 4000 frames, was about 50 hours. The advantage of this method is demonstrated in Table 3.1 where much higher journey-time sampling rates are shown from the use of cine-photography, with an overall sampling rate of 83% in one survey. The average journey-time sampling rate of the cine-surveys was 63%.

### 3.2 ANALYSIS.

#### 3.2.1 Preliminary Analysis.

Preliminary analysis of the data was carried out to determine the means, standard deviations and the ranges of the various parameters. It was found helpful to plot the simultaneous occurrence of flows and journey-times on a graph with elapsed time as the ordinate. For 1 minute sampling intervals the scatter was substantial; for example, for the data collected on 29/10/73 the range was between 2 and 16 veh/min. In order to reduce the scatter and reveal trends in the variables during periods of observation moving-average flows were calculated using computer programs which permitted a choice of the period of the moving-average. The effects on the relationships of choosing different periods will be discussed later in this thesis. This smoothing process is vital if underlying trends are not to be obscured by the scatter of the data. Whether the moving-average process should be carried out on the journey-time data is debatable, since the journey-times used were already averaged over the number of vehicles whose journey-times were measured in a particular sampling interval. For the first two surveys however, the journey-time sample

## FLOW/JOURNEY-TIME STUDIES

was relatively small, 17% and 24% respectively, corresponding to averages of 1.5 and 2.4 vehicles whose journey-times were measured in each one minute interval. The later surveys, using cine-photography, provided journey-time data which was much more reliable, and much less variation was observed between mean journey-times for adjacent intervals.

Figures 3.1 to 3.4 show elapsed time plots of 5 minute moving-average flows and 1 minute mean journey-times for the High Street and Southbury Road. They each indicate a different feature of the trends in the data. Fig. 3.1 reveals apparent periodicity with two major peaks in both journey-time and flow and a minor one in flow at approximately 30 minutes separation which may be caused by employees leaving different factories at particular times. Similar answers were obtained using moving-average flows calculated over a series of periods between two and ten minutes. Fig. 3.2 shows the data for Southbury Road (E/Bound) on 19/10/73 and indicates unexpected negative correlation between journey-time and flow for the first 50 minutes. Fig. 3.3 shows a steady increase in each of the variables in the morning peak in Southbury Road (E/Bound) with high values at the end of the period. Fig. 3.4 represents the data for Southbury Road (W/Bound) for the morning peak period on 5/12/73, showing a distinct peak in the mean journey-times, caused by the queue which developed at a petrol station.

### 3.2.2 Regression Analysis and Significant Correlations

The following linear regression analyses have been carried out:

1. Between journey-time and total flow  
including log transformation, power



## FLOW/JOURNEY-TIME STUDIES

curves and serial correlations.

2. Between journey-time and classified flows.
3. Between journey-time and total flow, turning flow, pedestrians crossing and total opposing flow.

### Linear Regression Analysis

The results of the regression analysis of journey-time on entering traffic flow in the High Street are shown in Table 3.2 and a scatter diagram for the results collected on 19/7/72 is shown in Fig. 3.5. During the off-peak period the correlation coefficients were not significantly different from zero whereas in the peak period significant correlation coefficients were obtained. The long-term mean flows were similar for both periods.

The data in all of the studies have been aggregated over successively longer sampling intervals and the journey-times and flows, expressed in vehicles/minute, were then subjected to further linear regression analysis. Tables 3.4 and 3.5 illustrate the effects of this aggregation on the correlation and regression coefficients. The slope regression coefficients can be seen to increase consistently with increasing periods of aggregation in most samples of peak period data. These changes of the coefficients were predicted by Wright (1971); they can be observed only if a significant degree of serial correlation exists between journey-time and flow for the longer periods of lag. Thus, these changes are evident in the data collected

## FLOW/JOURNEY-TIME STUDIES

in the High Street on 19/7/72 where journey-times were significantly correlated with flows occurring up to eight minutes earlier. They were not observed in the data collected in the off-peak period on 17/7/72 (see Table 3.3), where the slope regression coefficients can be seen to vary haphazardly with changes in the period of aggregation.

It is possible that random variations of flows and journey-times in small sampling intervals may be of such a scale as to hide small systematic changes in journey-time. The data collected in the High Street in 1972 is particularly prone to this effect because of the small samples of vehicles whose journey-times were measured. However, the data collected on 22/10/73 also demonstrates the progressive changes in the values of the correlation and regression coefficients as the data is aggregated and in this data the journey-time sample was 83%. Since such differences are evident between the coefficients for data aggregated over different sampling intervals within the same study, such differences are also likely to arise between the results of different studies using different sampling intervals even when other factors are similar. Thus the satisfactory comparison of the results of different studies may be difficult to achieve, whenever different sampling intervals have been used. A further difficulty may arise in studies where little or no control is exercised over the size of the sampling intervals, for example, in studies where the moving-observer method is used for measuring flows and journey-times and, in effect, short sampling intervals are used. In such studies any real variation in the relationships between different roads may be obscured by variations generated by differences in the size of sampling intervals.

## FLOW/JOURNEY-TIME STUDIES

The trends in the values of the slope regression coefficients observed in these studies are discussed by Wright (ibid), who considers the causes of these trends in some detail. The effects are asserted to arise as a result of two processes which he calls Bias Types 1 and 2. Bias Type 1 arises because the measures of journey-time or journey-speed in an interval are normally weighted according to the flow in that interval. When data are aggregated in adjacent intervals the mean flow is conventionally the arithmetic mean of the flows in the small sampling intervals, whereas the mean journey-time is the weighted mean of the journey-times in the small sampling intervals. Thus the overall mean journey-time in the longer sampling intervals is biased towards the mean journey-time of the intervals in which the highest flows occurred. Wright concluded that this effect may not be important and, in any case, it can be eliminated if unweighted mean journey-times are used. Bias Type 2 arises out of the serial and cross-correlations between the variables in adjacent intervals. In the present study the change in the slope regression coefficients as a result of aggregation has been found to be significant even when unweighted journey-times are used, as was previously shown by Wright. Thus the greatest effect of aggregation is to give rise to Bias Type 2, and this is shown in Table 3.2a where the coefficients for models using both weighted and unweighted samples of mean journey-time are compared. These results suggest that if two studies are being compared then the sizes of the sampling intervals should be equal. If different sampling intervals are used Bias Type 1 cannot be eliminated from the study having the larger sampling intervals and Bias Type 2 exists in both sets of results. Even if studies having equal sampling intervals are compared it is not evident that Bias Type 2 will have equal effects in both sets of results.

## FLOW/JOURNEY-TIME STUDIES

Linear regression analysis of the various samples of data collected on Southbury Road has also been carried out and these results are summarised in Table 3.4. The results for this road are much more difficult to assess. For the eastbound traffic flow significant linear relationships were obtained between flow and mean journey-time for the morning peak but not for the evening peak period. For the westbound flow significant relationships were obtained for the evening peak period but the results for the morning peak were influenced by a queue which developed at a petrol station. The data was collected during the 1973 fuel crisis and the queue developed about 8-30 am when the petrol station opened. The scatter diagram for the latter sample is shown in Fig 3.6 with a parabolic curve drawn through the points which reproduces a turning-back effect which has been postulated for the theoretical speed-flow relationship - see Smeed (1949). The turning-back effect in this case is caused by the reduction in capacity arising out of the congestion near the petrol station. This effect is even more pronounced in the eastbound data collected on 19/10/73, see Fig.3.7, and was probably caused by delays at traffic signals beyond the end of the survey section. These results tend to confirm the argument presented by Branston (1976) that such an effect will only be observed when the entry capacity exceeds the exit capacity for a particular link and then only if the sampling interval is small enough to detect relatively rapid changes in flow and journey-time. That is to say, when the sampling interval is much shorter than the period of time for which the entry flow exceeds capacity. Any phenomenon which relies for its observation on the selection of a particular survey point may not be representative of the behaviour over an extended section of road and this is perpetually one of the major difficulties for investigators of speed-flow

## FLOW/JOURNEY-TIME STUDIES

relationships since some investigators have chosen to collect their data at a point. However, since there will continue to be a need to provide some means of predicting traffic speeds in assignment models then speed-flow models derived from point observations are likely to be used. More reliable models might be obtained using the definitions of flow and speed defined by Edie (1963) - flow is the aggregate distance travelled by all vehicles passing through a space-time domain divided by the area of the domain and speed is the aggregate distance travelled divided by the aggregate time spent travelling. Thus the inaccuracies arising in speed-flow models derived from point measurements of flow might be avoided using the flow as defined by Edie, but the problems of data collection could be insurmountable in some circumstances.

If Branston's assertion is accepted then the journey-time/flow characteristics in a link will be governed by the point having the lowest capacity. In this case a significant relationship could be obtained but it may not be linear if the data has a wide range of values. The departure from linearity can be detected by examining the residuals derived from a linear regression model. In this study an examination of the residuals of the linear regression model indicates that curvilinear or two-regime models may be more appropriate, see Fig 3.8. Examples of models of these types will be discussed in a later chapter.

### 3.2.3 Multiple Regression Models.

Multiple regression analysis of the traffic flow, subdivided into vehicle classes, has been carried out on the data collected in the High Street on 19/7/72. The correlation matrix reveals that this subdivision tends to reduce the significance of any relationships

## FLOW/JOURNEY-TIME STUDIES

between vehicle flow and mean journey-time since only the flow of cars shows any significant correlation and the coefficient is less than that between total flow and journey-time. No significant multiple regression models were obtained using traffic flows subdivided into vehicle classes because of insufficient variations in the variables lorries, buses and motor-cycles.

In addition multiple regression analysis has been carried out using four "independent" variables, namely; total flow, number of pedestrians using the zebra crossing, numbers of vehicles turning at the intermediate junctions and total opposing flow. The correlation matrix is shown in Table 3.6, where the data has been aggregated over intervals of five minutes. Similar results were obtained with the data aggregated over other periods although, like the regression coefficients, the correlation coefficients display consistent trends as the data is aggregated over successively longer sampling intervals. The table reveals that the pedestrians crossing and the turning vehicles could each be included in a separate linear regression equation with mean journey-time having a level of significance similar to that between total flow and mean journey-time. The table also shows that these two variables and total flow are highly intercorrelated. The full range of possible multiple linear regression equations and their coefficients are given in Table 3.7. The equations marked with an asterisk have correlation coefficients which are significantly greater than zero at the 5% level. Because of the intercorrelations between the "independent" variables the regression coefficients of the same variables may vary considerably between different models. None of the multiple regression equations has a correlation coefficient which is significantly greater than the

most significant equation containing one independent variable. The multicollinearity could be expected at this time of day since the variables all reflect the greater activity in the peak period. A method of combining the data into an index of peak period activity will be discussed later.

#### 3.2.4 Analysis of Serial Correlation Relationships

Since Wright (ibid) asserts that Bias Type 2 arises as a result of the serial correlations between the variables, the relationships between flow and journey-time with various lags between the variables have also been studied. Lags from 0 to 9 minutes have been considered between the following variables:

1. Journey-time and lagged flow.
2. Journey-time and lagged journey-time.
3. Flow and lagged journey-time.
4. Flow and lagged flow.

The purpose of this part of the study was to determine for how long intercorrelations persisted and their importance. The results of the analysis of the data collected in the High Street on 19/7/72 are given in Table 3.8. Many significant correlations exist between journey-time and lagged flow but there are few between flow and lagged journey-time. The significant correlations may persist for up to 7 or 8 minutes. However, the results must be considered very carefully since the correlations observed may simply reflect time dependent variations. This is particularly true of the data collected on Southbury Road (eastbound) on 5/12/73 in the morning peak period where both long term mean flows and journey-times tended to increase

## FLOW/JOURNEY-TIME STUDIES

monotonically. Furthermore these results show that the intercorrelations between journey-time and lagged journey-time are highly significant. Similar results were obtained in all sets of peak period data. The cause of this high correlation can be explained by the lower variation of mean journey-time between adjacent intervals compared with off-peak periods and the persistence of the intercorrelations between journey-time and lagged flows up to 7 or 8 minutes. If the relationship between journey-time and lagged flow up to 7 minutes earlier is assumed to be causal then the mean journey-time in an interval and that in the next interval will be mutually correlated with flows up to 6 minutes earlier. Thus high intercorrelations of journey-times can be expected.

Another time-related phenomenon is also relevant to this section, namely the time-lags between peaks and troughs in the flows and the corresponding peaks and troughs in the mean journey-times. This is illustrated in Fig 3.1. The explanation of the cause of this phenomenon may also be related to the argument presented by Branston (1976), and discussed earlier in this thesis, that there is a lag between the decay of the departing flow from a section after the entering flow has reduced as queueing vehicles in a congested link disperse.

Wright (1971) produced a mathematical model to predict slope regression coefficients for data aggregated over  $n$  intervals from a set of measured lagged serial and cross covariances. The model is expressed as follows:



$$B_n = \frac{C_{fj}(0) + \sum_{s=1}^n M(s)(C_{fj}(s) + C_{jf}(s))}{C_{ff}(0) + \sum_{s=1}^n 2M(s)(C_{ff}(s))}$$

where  $B_n$  = predicted slope regression coefficient  
between journey-time and flow aggregated over  
 $n$  intervals.

$$M(s) = \frac{n-s}{s}$$

$f$  = subscript for flow

$j$  = subscript for journey-time

$C_{fj}(s)$  = cross-correlation between journey-time  
and flow lagged by  $s$  intervals.

$C_{jf}(s)$  = cross-correlation between flow and  
journey-time lagged by  $s$  intervals.

The model reveals the source and effects of Bias Type 2 as discussed by Wright (ibid). In deriving the model it was assumed that the variables were generated by a stationary process and that the covariances became progressively smaller with increasing values of  $s$ . An identical argument can be used to develop a similar model for the

## FLOW/JOURNEY-TIME STUDIES

correlation coefficient of the aggregated relationship. A number of practical difficulties occurred when using these models in this study. Because of the sampling method used, in some intervals no measurements of journey-time were available. Also, because of the approximations used, it was possible for some of the lagged serial covariances to become negative and hence the square roots are unreal.

These models have been used to predict slope regression and correlation coefficients for successively longer sampling intervals. In general, for data collected at peak periods, the predicted coefficients tend to increase to some maximum value as the covariances corresponding to greater lags are included in the models. Beyond the maxima the slope regression coefficients tend to decrease and the correlation coefficients vary haphazardly about the maximum. The results for Southbury Road (westbound) on 19/10/73 are shown in Table 3.9. where the predicted coefficients are compared with the corresponding coefficients calculated using the aggregated data and the results agree closely. In Table 3.10 it can be seen that as the covariances corresponding to successively greater lags are included in the model, the maximum slope occurs with different numbers of lagged covariances included in each sample. These comments are generally true for all of the peak period samples. No apparent relationship could be observed between the number of lagged covariances included in the model which produced the maximum slope regression coefficient and the initial value of the slope regression coefficient of the 1-minute samples. There is some indication that this is the result of some time-related phenomenon in the data since, in the case of the sample collected in the High Street on 22/10/73 where both flows and mean

## FLOW/JOURNEY-TIME STUDIES

journey-times tended to increase continuously during the period, the maximum values of the coefficients could not be predicted.

If these predictor models are to be useful then it is evident that the same maximum values of the coefficients should be obtained no matter what size of sampling interval is used in collecting the original data. In order to test this hypothesis the data collected in the High Street on 19/7/72 was aggregated into samples based on successively larger sampling intervals and Wright's models were used to determine the maximum values of the coefficients. Table 3.11 shows the results of this analysis. The results show close agreement on both the value of the maximum slope and the optimum period of lagged covariances to be included in the models. The predicted correlation coefficients are considered to be much less reliable for the reasons given earlier. The results of these tests of Wright's model for predicting slope regression coefficients suggest practical methods of comparing results between studies having different sampling intervals. There appear to be two alternatives; (a) aggregate the data so that the new sampling intervals in each study are of equal size, or (b) use Wright's model to predict the maximum slope regression coefficients for each study. The results of this study indicate that (b) would be the preferred method since (a) may require a prohibitively large number of samples.

### 3.2.5 Analysis of Pseudo-Concentration Relationships.

Concentration is usually very difficult to measure directly on roads longer than a few hundred yards. Consequently a number of researchers have resorted to indirect methods using the relationship  $Q = KV$ , or the related equation  $QT/L = K$ ,

where  $T$  is journey-time over the distance  $L$ .

## FLOW/JOURNEY-TIME STUDIES

Concentration derived by this means will be called pseudo-concentration in this report. The relationship between journey-time and pseudo-concentration has been investigated for all samples of data collected in this study and usually high correlation coefficients were obtained. Other researchers, Guerin (1955) and Duncan (1979) have also reported high correlation coefficients with these variables. Duncan argues that the translation of speed/concentration relationships into the speed/flow form is not valid because a wide scatter of the low speed data invariably occurs. Furthermore such relationships cannot be used for predictive purposes directly since the dependent variable occurs on both sides of the expressions. For example, an equation relating journey-time and pseudo-concentration may be expressed by:

$$T = A + B\left(\frac{QT}{L}\right)$$

By a simple rearrangement  $T$  may be obtained, thus

$$T = \frac{A}{1 - \frac{BQ}{L}}$$

However, the correlation coefficient is usually lower for the rearranged equation and in many instances not statistically different from zero. The correlation coefficients of the regression equations between journey-time and pseudo-concentration obtained in this study

## FLOW/JOURNEY-TIME STUDIES

are compared with those between the observed and predicted journey-times derived from the above equation in Table 3.12. The table reveals the inadequacy of these models for predictive purposes and it is doubtful whether they should ever be used.

## CHAPTER 4

### TWO-REGIME AND NON-LINEAR MODELS

#### 4.1 DISCONTINUOUS MODELS.

The term "discontinuous" is used in this study to describe a two-regime model of the type shown in Fig. 4.1. One of the earliest studies of traffic flow and speed was reported by Wardrop (1952). A feature of his results was the use of a discontinuous model to express the variation of the parameters shown as follows:

$$V = 31 - \frac{Q + 430}{3(w - 6)} \quad \text{or } 24\text{mph}$$

whichever is less,

where  $V$  = running speed (mph)

$Q$  = traffic flow (vph)

$w$  = road width (feet)

The results were also represented graphically, showing more clearly the nature of the relationship for typical road widths (see Fig. 2.1). For most roads the relationship implies that there is a critical flow above which changes in running speed are correlated with changes in flow, and below which these variables are independent. Similar types of models have been postulated by Dick (1966), Smock (1962) and Davidson (1966), the latter two models differing from Wardrop's in the form of the assumed relationship above the critical flow.

## TWO-REGIME AND NON-LINEAR MODELS

The data in the present study has been analysed to determine to what extent a discontinuous model might be applicable. The reliability of the results is difficult to assess since the usual tests of significance are not valid. The computer program written for this analysis carries out a step-by-step process; it chooses an initial value of the critical flow, calculates a number of statistics and then increments this value in steps of 0.5 veh/min. At each step it calculates the mean journey-time for all intervals in which the flow is less than the critical value. For flows greater than the critical value, it calculates the equation of the line through the point of mean flow and mean journey-time which gives the same value of journey-time at the critical flow as the average for flows less than the critical flow. The principle is illustrated in Fig 4.1. The statistic used to compare the results for the different values of the critical flow is the combined sum of (a) the total sum of squares of the deviations of journey-times for flows less than the critical flow, and (b) the residual sum of squares of the journey-times about the sloping line for flows greater than the critical flow. The program can also be used with journey-speeds instead of journey-times.

The results of this analysis for a representative set of data are given in Table 4.1. In this example the data has been aggregated and the values used in the model are the 5 minute averages of both flow and journey-time. The table shows that the minimum combined sum of squares of journey-times is obtained if the value of the critical flow is assumed to occur at a flow of 8.50 veh/min. and the value of the combined sum of squares is  $17.61 \text{ min.}^2$ . A linear regression analysis of this data gives a value of the residual sum of squares of  $18.50 \text{ min.}^2$ . This example demonstrates that the discontinuous model

## TWO-REGIME AND NON-LINEAR MODELS

explains a larger proportion of the variance of journey-time than a linear regression model. At this stage it is not clear whether this is a significant improvement in the explanatory power of the models.

The discontinuous model has been used on the same data aggregated in a different way using moving-average flows, instead of simple average flows, and one-minute mean journey-times. The purpose was to determine if the use of moving-averages affected the apparent position of the point of discontinuity or revealed further evidence of the relevance of a discontinuous model. In this case the minimum combined sum of squares was  $14.59 \text{ min.}^2$  and this occurred when the point of discontinuity was located at a mean flow of 9.50 veh/min. A linear regression analysis of the same data gave a residual sum of squares of  $20.03 \text{ min.}^2$  which is considerably greater than the value given by the discontinuous model. The relationship being tested is that between the mean journey-time in a one-minute interval and the mean flow averaged over the period up to ten minutes earlier. For this data this is clearly a significant relationship since for the least-squares model the F ratio is 96.5. However, since adjacent values of moving-average flow are highly correlated there are theoretical objections to using moving-averages in predictive models.

The effects on regression and correlation coefficients of aggregating data over different numbers of sampling intervals has been discussed earlier when considering least-squares regression models and similar effects were observed when using the discontinuous model. Table 4.2 shows the results of aggregation up to intervals of ten minutes using simple averages. The results show almost consistent reductions in the relative values of the mean square error as the data is aggregated over successively longer periods and consistent increases in the slope



## TWO-REGIME AND NON-LINEAR MODELS

of the relationships for intervals when the flow is greater than the critical flow. This consistent pattern in the results was discussed in the last chapter and no further comment is needed. It can also be seen that the mean square error of the discontinuous model is less than that of the "least squares" model.

The discontinuous model has been used on the combined data to investigate journey-speed flow models and the results are compared with those obtained by least-squares regression analysis of the same data in Table 4.3. In this case the mean square errors do not differ by very much except for periods of aggregation of 4 and 10 minutes. Thus the discontinuous model does not succeed in explaining more of the variance of journey-speed than the more commonly used linear regression model. However, there is no reason to believe that this will generally be the result.

There appears to be no generalised statistic which can be used to determine whether the discontinuous model fits the journey-time/flow data better than the journey-speed/flow data since the F ratio is not valid in this case. However, for the above data a more tortuous argument might be used. For the data aggregated over 5 minutes the journey-time/flow and journey-speed/flow models gave F ratios of 18.8 and 14.9 respectively when using the least-squares method. Since for journey-time data the discontinuous model gives a better fit than the linear regression model, and for the journey-speed data the discontinuous model gives no better fit than the linear regression model, hence the discontinuous model can be expected to give a better fit to journey-time data than to journey-speeds.

## TWO-REGIME AND NON-LINEAR MODELS

### 4.2 TWO-REGIME MODELS.

In Chapter 3 it was stated that an examination of the residuals of a linear regression model showed that a non-linear model could be expected to provide more satisfactory predictions. Furthermore a model which provides a gradual progression from an independent to a dependent zone would be expected to be more representative of traffic behaviour than the discontinuous linear model. One such model which has been tested uses an undefined power law for flows greater than the critical flow. It is probably more correct to call this a two-regime model; however, there are a number of similarities with the discontinuous model which will become evident as the model is discussed. For flows greater than the critical flow the model may be expressed as follows:

$$(T - T_o) = V_o(Q - Q_k)^B$$

where  $T_o$  = mean journey-time for flows less than or  
equal to  $Q_k$

$Q_k$  = critical flow

$V_o$  and  $B$  are empirical constants.

Let

$$y = (T - T_o)$$

and

$$x = (Q - Q_k)$$

then the above equation can be transformed to

$$\log y = \log V_o + B \log x$$

which is now in a suitable form for linear regression,

$$Y = A + BX$$

where  $Y = \log y$

$X = \log x$

## TWO-REGIME AND NON-LINEAR MODELS

A = an empirical constant.

This model has been tested on a number of sets of data and the results for the High Street are given in Table 4.4.

The mean square errors in this table can be compared with those in Table 4.2 and this shows that the power law gives a better fit than the linear relationship. However, there are a number of problems with this model. Firstly,  $y$  cannot be negative since the logs of negative numbers cannot be obtained; hence if the measured journey-times were less than  $T_0$  in intervals when the flow was greater than the critical flow, then these results were omitted from the calculation. Thus a number of samples were left out having flows just a little larger than  $Q_k$ , thereby influencing the calculated value of the slope. The values of  $B$  were almost always numerically less than 1.0, so the curves would be steepest at flows close to  $Q_k$  and flatter at higher flows. The optimum values of  $B$  are close to zero in most cases, thus wrongly suggesting that a single average value of journey-time would be appropriate for all intervals in which the flow is greater than  $Q_k$ . Altogether these results suggest that the power law is unsuitable.

### 4.2 DAVIDSON'S MODEL.

The derivation of a flow-travel time relationship based upon queueing theory is described by Davidson (1966), which he justifies by arguing that journey-time would be expected to vary with flow according to a monotonic function and approach infinity as the level of flow approached the saturation flow. Assuming random arrivals and service in a length of road then the ratio of delay ( $w$ ) to service time ( $t$ ) is

## TWO-REGIME AND NON-LINEAR MODELS

given by:

$$\frac{w}{t} = \frac{c}{(1 - c)}$$

where  $c$  = ratio of arrival flow to saturation flow

and  $c = Q/S$

where  $Q$  = traffic flow

$S$  = saturation flow

Davidson argues that traffic flowing along a road is subjected to a series of queueing situations depending upon the type of road and the frequency of delay-causing occurrences along its length, and so he introduces an additional parameter  $J$  to account for these occurrences. Thus the equation is modified to:

$$\frac{w}{t} = \frac{Jc}{(1 - c)}$$

Furthermore, the total travel time  $T$  is made up of two terms namely, delay and service time or travel time at zero flow, giving:

$$T = t + w$$

$$= t + \frac{tJc}{(1 - c)}$$

$$= \frac{t(1 - c(1 - J))}{(1 - c)}$$

There are a number of practical difficulties in using this equation since, in general,  $t$ ,  $J$  and  $S$  are not known for a particular road. In Davidson's paper he describes an optimising technique, used on a small sample of data, which minimises the mean square error between the predicted and observed travel times over a range of traffic flows. He suggests that  $J$  is a measure of road service quality; roads which have a low value can carry traffic flows approaching saturation with

## TWO-REGIME AND NON-LINEAR MODELS

very little delay whereas roads having higher J values may have high delays at flows much less than saturation. These results are illustrated in Fig. 4.2.

Menon et al (1974) considered some of the problems associated with the use of this formula, such as the conditions under which zero flow travel time should be measured and the effects of delays at signal-controlled intersections, which were not subtracted from the travel times in Davidson's results. They also attempted to determine the values of typical zero flow travel times for different types of areas (commercial or residential) and the factors which influence the J values. No close relationships were obtained due to the high variance of the observed journey-times.

Taylor (1977) describes a method which enables simultaneous least-squares estimates of the parameters t, S and J to be determined from observed data. In general there will be an error between the observed travel time and the predicted travel time at a specific traffic flow. Let U be the sum of the square errors for all of the values of observed journey-time, then the following derivatives

$$\frac{dU}{dS}, \frac{dU}{dJ} \text{ and } \frac{dU}{dt}$$

will provide a series of simultaneous equations which permit S, J and t to be obtained from certain functions of T and Q. He concludes that t and S can be calculated with reasonable accuracy but J is subject to larger proportional variations. Akcelik (1978) compared Davidson's travel time function with a hyperbolic cost function derived by Mosher (1963) and concluded that these were equivalent. Mosher uses a quality of service parameter (m) which is related to Davidson's factor

## TWO-REGIME AND NON-LINEAR MODELS

as follows:

$$J = 1 - m$$

He showed that  $m$  was the ratio of the function's asymptote to free flow travel time and that this determines the flatness of the function at low values of flow. The asymptote is the value of  $T/t$  which Davidson's function approaches as  $c$  tends towards minus infinity. Akcelik described a modified version of the function which allows finite travel times to be derived for oversaturated flow conditions in order to permit realistic traffic assignments to be carried out. Davidson's function has been used, in its original form, in order to analyse the data collected in this study using the minimum mean square error of the predicted journey-times to determine the optimum values of  $S, J$  and  $t$ . Taylor's equations were not used since they were found to be far more difficult to compute than the approximate method which Davidson used. The approximate method can be described as follows; using Davidson's original formula, the predicted journey-time is given by:

$$T = t(1 + \frac{Jc}{(1 - c)})$$

Now the sum of the square errors  $U$  is given by:

$$U = \sum (T_i - T)^2$$

where the subscript  $i$  refers to the  $i$ th interval.

Thus

$$\frac{dU}{dJ} = -2t \sum (T_i - T) \frac{dT}{dJ}$$

## TWO-REGIME AND NON-LINEAR MODELS

$$= - 2t \sum (T_i - t (1 + \frac{JC_i}{1 - C_i})) (\frac{C_i}{1 - C_i})$$

$$= 0$$

Now substituting for  $c$  and solving for  $J$  gives:

$$J = \frac{\sum (T_i - t) (\frac{Q_i}{S - Q_i})}{t \sum (\frac{Q_i}{S - Q_i})^2}$$

This equation has been used in this study to determine the values of  $J$  for a series of values of  $S$  and  $t$ , which were altered incrementally and the optimum value of  $J$  determined by inspection of the computed value of the mean square error.

The results of this analysis on the High Street data are given in Table 4.5 and compared with other models in Fig 4.3. The mean square errors of the predicted journey-times may be compared with those given in Table 4.2 for the linear and the discontinuous models and it can be seen that they tend to be lower than those given by the linear model but higher than those given by the discontinuous models. In addition it can be seen that there are systematic changes in the optimum values of the parameters as a result of aggregation of the data into successively larger sampling intervals.

## TWO-REGIME AND NON-LINEAR MODELS

The advantage of the Davidson model is that the same relationship is used throughout the whole range of flows and it has an underlying theoretical basis, whereas the models used earlier were entirely empirical and their use can only be fully justified within the range of values of flow and journey-time which were used for their derivation. The disadvantage of the Davidson model is that accurate values of the parameters may be difficult to obtain since they have not been widely investigated. One of these parameters, off-peak travel time, is the subject studied in later chapters of this thesis. Menon et al (1974) discuss a series of experiments to determine the nature of  $J$  and its dependence upon environmental and other roadside characteristics. The results of their study were inconclusive, largely because of the narrow range of traffic flows over which the model was tested and the high variance of the parameters measured. They did not find it possible to determine the influence of particular environmental factors on the parameters  $T_0$  and  $J$ . To date no further progress has been reported on the accurate measurement of the factors affecting  $J$  and until this is available the model is unlikely to be widely used. However, saturation flows are probably more easily measured and it is possible to account for the effects of aggregation on saturation flows using a probability model. Referring to Table 4.5 it can be seen that there are systematic trends in the values of saturation flow and at this stage it is not evident whether these changes are caused by the same statistical mechanisms as those which affect the regression coefficients. However, the following theoretical argument can be used to derive a model to predict the theoretical values of saturation flow for different periods of aggregation.



## TWO-REGIME AND NON-LINEAR MODELS

Assume that the numbers of vehicles passing a point on a road are generated by a stationary process such that, throughout the period of observation, the long-term mean flow remains constant.

Define  $P_i$  = probability of  $i$  arrivals during an interval, and

$$Q_k = \sum_{i=0}^k P_i \quad ; \text{Prob } (0 \leq i \leq K)$$

In a series of  $n$  intervals, the probability of  $k$  arrivals in any one interval and less than  $k$  in any of the remainder is

$$n P_k (Q_{k-1})^{n-1}$$

assuming that the numbers of arrivals in each interval are independent. In general, the probability of exactly  $k$  arrivals in each of  $r$  intervals and less than  $k$  in the remaining  $n-r$  intervals is given by:

$$P_k = \frac{n!}{r! (n-r)!} P_k^r (Q_{k-1})^{n-r}$$

Hence in  $n$  intervals the probability of  $k$  arrivals in any number of intervals and no more than  $k$  is:

$$\begin{aligned} &= n P_k (Q_{k-1})^{n-1} + \frac{n(n-1)}{2!} P_k^2 (Q_{k-1})^{n-2} + \dots \\ &+ \frac{n!}{r! (n-r)!} P_k^r (Q_{k-1})^{n-r} + \dots + P_k^n \\ &= Q_k^n - Q_{k-1}^n \end{aligned}$$

## TWO-REGIME AND NON-LINEAR MODELS

Since this is the probability of  $k$  arrivals at least once in  $n$  intervals and no more than  $k$  arrivals in the others, it is also the probability that  $k$  is the maximum number of arrivals in any interval in the series of  $n$  intervals. It follows that the expected maximum number of arrivals is given by:

$$\text{Exp } (k_{\max}) = \sum_0^{\infty} k (Q_k^n - Q_{k-1}^n)$$

The probabilities may be derived from a Poisson or a normal distribution to permit theoretical values of the expected maximum value of  $k$  to be determined. Furthermore, by selecting appropriate values of  $n$  and probability the effects of different lengths of sampling intervals may be investigated.

The model has been used to predict saturation flows along the High Street based upon the total data. Both Poisson and normal probability functions have been used and the results are given in Table 4.6. Although the Poisson distribution has been used in this model it is unlikely to provide satisfactory estimates of saturation flow because, as can be seen from the table, the ratio of variance to mean is much greater than 1.0 for all periods of aggregation. In the table the long-term saturation flow has been assumed to be equal to the mean flow over the period of study for the Poisson distribution and the column labelled "normal distribution (1)". It can be seen that the values of the corresponding saturation flows are lower than those predicted by Davidson's model. Clearly the accuracy of these values depends upon a satisfactory estimate of the long-term saturation flow. The saturation flows in the column labelled "normal distribution (2)" are based upon a long-term saturation flow assumed to be 1.1 times the

## TWO-REGIME AND NON-LINEAR MODELS

maximum observed ten-minute flow measured in the peak period, that is,  $1.1 \times 12.0 = 13.2$  veh/min, and the observed variance. As would be expected, this latter set of predictions gives closer agreement with the values predicted by Davidson's model.

It is evident that the model succeeds in predicting consistent changes in the saturation flows with changes in the sampling intervals. The probability argument used in the peak-flow prediction model provides a satisfactory explanation of the reason for these changes. Since high flows are only likely to occur for relatively short periods then the average saturation flows must become progressively smaller for the longer periods of aggregation. This phenomenon has numerous counterparts in applied science, for example, the changes of saturation flow are directly analogous to the changes of maximum rainfall rates in storms of successively longer periods as reported by Bilham (1935). Another example is the prediction of maximum wind loadings on structures which is discussed by Davenport (1964).

The arguments used in deriving the model for predicting saturation flows for different durations can also be used to account for the discrepancy between the theoretical capacity and the observed capacity of a single lane of traffic flow. The theoretical capacity, which is given as 2750 vehicles/hour by Smeed (1948), was derived from the results of investigations of space-headways in controlled and open-road conditions. The empirical equation for modal headway as a function of speed was expressed by:

$$h = 17.5 + 1.17 V + .008 V^2$$

where  $h$  = modal headway (feet)

$V$  = vehicle speed (mph)

## TWO-REGIME AND NON-LINEAR MODELS

Since space headway is the reciprocal of concentration this equation was substituted in the fundamental relationship of traffic flow to permit the following equation to be obtained:

$$Q = \frac{5280 V}{17.5 + 1.17V + .008V^2}$$

where Q = traffic flow (veh/hour)

Almost invariably observed maximum flows in saturated conditions are much less than the theoretical value if the observations are carried out for a period of one hour. The reason is that the headway equation was derived for short time intervals since only the headways corresponding to vehicles which were obviously maintaining the minimum safe headways were included in the equation. It is not evident to which time sampling interval the theoretical capacity is related since some of the short space-headways must have occurred singly (that is, a short headway was measured between an isolated pair of vehicles) and the remainder in platoons of varying lengths. However, it is clear that the effective period is much less than one hour.

## CHAPTER 5

### MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

#### 5.1 INTRODUCTION.

A significant difficulty in using speed-flow relationships in towns is that extraneous factors, such as pedestrian crossings, parked vehicles and turning vehicles, may have as significant an effect on journey-speed as changes in flow. Wardrop (1952) recognised this in the formulae for journey-speeds on London streets. A number of other studies have included some of these factors in order to allow for the effects in the resulting regression equations. Often these extraneous variables may be intercorrelated with each other and with traffic flows since peak flows of traffic, turning vehicles and pedestrians are likely to coincide. Such variables may be rejected from multiple regression models if the commonly accepted tests of significance are used and if the researcher chooses to retain these variables the increase in correlation coefficient will be small. The same kind of problem arises if traffic flows are separated into different vehicle classes as discussed in Chapter 3. Sometimes researchers have resorted to the use of empirical indices in order to retain the effects of several variables - the use of pcu factors is an example. Otherwise, unless very large samples are obtained, a large proportion of the variables are likely to be rejected if traffic is subdivided into separate vehicle classes. The intercorrelation between the

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

"independent" variables may indicate interdependence, not necessarily directly causal, or an apparent correlation due to chance, and so the results need to be treated with caution. Where these variables are considered to be important, and must be retained, other methods of analysis should be used which can take account of these intercorrelations. Multivariate methods of analysis are used for dealing with this kind of data.

Apart from multiple linear regression analysis some of the more widely used techniques of multivariate analysis are principal component analysis, factor analysis and canonical correlation analysis. These methods have been widely used in psychology since they are capable of revealing evidence of latent factors of human behaviour which may be partially hidden in the results of various psychological tests; examples are discussed by Lawley (1971) and Thurstone (1965). They have also been used extensively in geology by Joreskog (1967). In the transport field they have been used in transportation planning by Fredland (1976), analyses of traffic speeds by Oppenlander (1963) and Wortman (1965), analyses of traffic flows by Gipps (1977), driver attitudes by Goldstein (1958) and road accidents by Versace (1960). In view of the multivariate nature of transport problems, and their relationships to human behaviour, one might expect a more widespread use of multivariate methods; however, with the exception of linear regression analysis this is just not the case. This is especially surprising when one realises that some of these techniques are capable of dealing with the problems of intercorrelation which occur frequently in transport studies.

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

In this chapter, after a brief introduction to the concepts of principal component analysis and common factor analysis, the use of these multivariate techniques in the analysis of journey-time data will be discussed. The theoretical principles of principal component analysis and factor analysis are described in more detail in Appendix E to this thesis.

Factor analysis is intended to study the intercorrelations which may exist between a number of variables and to examine the relationships which such correlations imply. The various methods attempt to identify hypothetical variables (called factors) which contain the essential information of the variations in the observed variables. The factors are linear functions of the original variables and they are so constructed that the overall complexity of the data is reduced by taking account of the inherent interdependence of the original variables. Usually the methods set out to find a smaller number of hypothetical factors which account for most of the information contained in the original variables. Furthermore, by comparison of the inter-relationships between the original variables and the factors, it may be possible to identify some common characteristics which these factors represent and to generate a better understanding of the phenomena being investigated.

Some authors, including Kendall (1975), Cooley and Lohnes (1971), regard principal component analysis as a separate method of analysis from common factor analysis and others, Thurstone (1965), Harman (1967), Joreskog et al (1976), consider it to be one method of factor analysis. The latter view is held because in a number of factor analysis techniques a principal component analysis is first carried out in order to provide data for the subsequent stages.

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

However, each of these methods is based upon a different model of the relationships between the factors and the original variables although the mathematical techniques used to derive the factors are similar in both cases.

### 5.1.1 PRINCIPAL COMPONENT ANALYSIS.

Principal component analysis attempts to simplify the data by finding a number of new variables, called principal components, which are linear combinations of the original variables but are themselves uncorrelated. Supposing that observations of  $p$  variables  $x_1 \dots x_p$  have been made on  $N$  individuals. The method of analysis tries to find  $p$  new variables  $e_i$  which are linear combinations of the  $x$ 's and are mutually orthogonal. Thus:

$$e_i = l_{i1}x_1 + l_{i2}x_2 + \dots + l_{ip}x_p$$

Almost invariably, the analysis first normalises the  $x$ 's before continuing with the remainder of the process, using the following expression:

$$z_i = \frac{x_i - \bar{x}}{S_x}$$

where  $\bar{x}$  = mean value of variable  $x$ ,

$S_x$  = standard deviation of  $x$ .

The functions  $e_1, e_2$  etc are the principal components and these are usually listed in an order which indicates the decreasing proportion of variance in the original data which is accounted for by each component. Furthermore, since the variables are normalised the total variance will be equal to the number of variables - for the case



## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

mentioned above the total will be  $p$ .\*

Using a geometric representation the original data can be imagined as  $N$  points in a  $p$  dimensional space. The principal component process can be imagined as redefining the axes such that they correspond to the principal components. The axis corresponding to the first principal component minimises the sum of squares of the distances of the  $N$  points from that axis. Similarly the axis corresponding to the second principal component minimises the sum of squares of the distances of the  $N$  points from that axis subject to the condition that it is orthogonal to the first axis, and so on.

### 5.1.2 COMMON FACTOR ANALYSIS

The factor analysis procedure to be discussed in this section has been given a number of different names such as "multiple factor analysis" by Harman (1967), "true factor analysis" by Joreskog et al (1976), "common factor analysis" by Taylor (1977) and "general factor analysis" by Nie et al (1975). The term "common factor analysis" will be used in this text since it describes most succinctly the mathematical objective of analysing the common factors which appear to influence the variables. The procedure is based upon ideas first suggested by Spearman in 1904, when he noticed systematic variations in the matrix of examination scores of a class of school children.

\*The analysis is carried out by solving the characteristic equation of the correlation matrix, thus each principal component is associated with an eigenvalue. The amount of the variance in the original data which is accounted for by each component is equal to the corresponding eigenvalue since the variables are normalised.

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

The following factor analysis procedure is intended to determine a set of factors which account for the maximum amount of common variability in the data. By common variability is meant the amount of variance of each variable which can be explained by its intercorrelations with the other variables. The basic model explicitly allows for both common and unique variance influencing the values of the measured variables, whereas in principal component analysis all of the variance is accounted for by the principal components. In addition, in common factor analysis the factors accounting for the common variance are fewer in number than the number of measured variables. This model can be represented as follows:

$$z_{ij} = c_{j1}h_{i1} + c_{j2}h_{i2} + \dots + c_{jp}h_{ip} + d_j u_{ij}$$

or

$$z_{ij} = \sum_{k=1}^p c_{jk}h_{ik} + d_j u_{ij}$$

where  $z_{ij}$  = the normalised value of the  $j$ th variable  
in the  $i$ th sample,

$h_{ik}$  = the  $i$ th value of the  $k$ th common factor,

$u_{ij}$  = the  $i$ th value of the unique (non-shared)  
factor associated with variable  $j$ ,

$c_{jk}$  = the weight assigned to the dependence of  
the  $j$ th variable on the  $k$ th common factor,

$d_j$  = the weight assigned to the  $j$ th unique  
factor.

Thus the weights are similar to regression coefficients and determine the observed values of the variables.

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

The model appears to be much less parsimonious than the principal component model since it involves  $p$  common factors and  $n$  unique factors. However, the unique factors are solely intended to permit the unique variance to be eliminated from the relationships between the original variables and the analysis concentrates on the common factor portion of the variables. The model represents the common proportion of each variable by the factors  $h_k$  and the unique portion by the unique variable  $u_j$ . Thus each variable may be regarded as having an element particular to itself and separate from the common causal phenomena; alternatively the residuals may be regarded as unique effects of the measured variables in contrast to the common effects of the common factors. The proportions of the variation of each variable attributable to each part are called uniqueness and communality respectively.

The initial solutions of principal component analysis or factor analysis are sometimes called direct solutions. Invariably in the former method the first principal component has high loadings on all of the variables and the remaining components tend to have fewer high loadings and to be bipolar; that is, they include both positive and negative loadings. Furthermore, since direct solutions always display such features their use in factor interpretation is limited, since the patterns occur as much as a result of the mathematical process as of the variations in the phenomena being studied. Such components are difficult to interpret since the wide range of factor loadings may not reveal the nature of the components. Consequently it has been argued, see Guilford (1968), that principal components do not reveal any structural feature in the data. Thus, in order to aid interpretation of the factors or to test a hypothesis, further solutions must be

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

obtained by means of rotation of the direct solutions, that is, rotation of the axes of the components. Various methods of rotation are used to help the interpretation of the factors by producing a "simple structure". Simple structure is defined as a factor pattern consisting of high values of the factor loadings on some of the variables and low values on the others since the important variables can be immediately identified. The principles of factor rotation are further explained in Appendix E and a diagram is presented which illustrates the process of rotation.

These methods will now be applied to the data collected in the High Street in 1972.

### 5.2 DEVELOPMENT OF FACTOR MODELS.

Data collected in the High Street on 17/7/72 and 19/7/72 has been analysed using these multivariate techniques and the correlation matrices are given in Tables 5.1 and 5.2 for data averaged over 5 minute intervals. For the data collected on 17/7/72 (off-peak) none of the correlation coefficients is significantly different from zero. For the data collected on 19/7/72 (peak period) the coefficients marked with an asterisk are significantly different from zero at the 5% level. It can be seen that for both sets of data some of the inter-correlations between "independent" variables are greater than those between the dependent and "independent" variables. For the peak data each of the "independent" variables has some significant intercorrelation with some other "independent" variable. There is no reason to assume that this is a chance result since it is common experience that not only traffic flows but also pedestrian movements and numbers of turning vehicles are all likely to increase and decay

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

more or less coincidentally at peak times. Thus there is evidence of the existence of factors whose variation at such times may influence these variables to some extent and which are only measurable by their effects on these variables. Principal component and factor analysis are intended to deal with the problem of revealing such latent factors.

A principal component analysis was carried out on both sets of data and the results are given in Table 5.3. For the peak period two factors having eigenvalues greater than 1.0 accounted for 89% of the variability of the data. However, for the off-peak period the eigenvalues are more closely clustered about the mean value of 1.0. If the variables were independent each eigenvalue would be 1.0, but for perfectly correlated variables the first principal component would have an eigenvalue of 4.0 (equal to the total number of variables included in the analysis) and the remainder would be zero. For the peak data the factor loadings of the first two principal components and the corresponding communalities are listed in Table 5.4. The communalities show the proportion of the variance of each individual variable which is accounted for by the common factors. Inspecting the factor loadings it can be seen that variables X1, X2 and X3 are largely dependent upon F1, and X4 is mostly dependent upon F2.

The principal components were then subjected to a varimax rotation (see Appendix E) to produce new positions of the axes of the orthogonal components and the results are plotted in Fig 5.1. Thus the interdependence of the factors and their associated variables becomes more evident. Factor F1, because it is positively correlated with variables X1, X2 and X3, will increase with increases in these variables and so it could be a good indicator of peak period activity.

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

Factor F2 is highly dependent upon the opposing flow X4 and is negatively correlated with pedestrian flow, thus reflecting the negative correlation between these two variables. A difficult question to answer about these results is why the opposing flow was not positively correlated with the other variables and, in particular X3, since the number of turning vehicles might be expected to be approximately equal for the two directions at this location.

The effects of aggregation of data over successively longer sampling intervals on the principal components has also been investigated. Since the factor loadings are the correlation coefficients between the variables and the corresponding factors it is not surprising that similar effects to those discussed in earlier chapters are observed in the results. These effects are demonstrated in the results given in Table 5.5 for the peak period data where the eigenvalues of the four principal components are listed for successively longer sampling intervals. The values for the longest intervals are very susceptible to sampling errors due to the small sample size. Clearly there is evidence in this table of consistent variations in the eigenvalues comparable to those variations observed in the correlation and regression coefficients discussed in Chapters 3 and 4. It seems likely, but by no means certain, that these eigenvalues would tend to converge on some specific values, as observed with the other statistics.

When using principal component analysis the most difficult task is to attach some meaning to the various isolated factors. Such analysis usually enables researchers to isolate interrelationships between some of the variables and from their knowledge of the nature of these variables they may be able to identify the underlying factors. If

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

many variables are being studied large volumes of data are generally needed in order to obtain reliable results. In this study two factors have been identified. The first factor has high positive factor loadings on traffic flow, number of crossing pedestrians and number of turning vehicles, and consequently high values of this factor will be associated with busy traffic conditions. Such a factor might prove useful as an index of peak period activity replacing traffic flow particularly where tidal flows occur. There does not seem to be any simple explanation for the occurrence of the second principal component.

### 5.3 FACTORS AND JOURNEY-TIMES.

A number of authors, including Kendall (1975) and Jolliffe (1973), have discussed the use of factors as independent variables in multiple regression analysis. Kendall postulates a model of the type:

$$G_j = \sum_{i=1}^{All i} a_i f_{s_{ij}}$$

where  $G_j$  = the  $j$ th value of the normalised dependent variable  $Y$ ,

$$G_j = \frac{Y_j - \bar{Y}}{SY}$$

$Y$  = mean value of  $Y$

$SY$  = standard deviation of  $Y$

$e_j$  = random error (zero mean)

$f_{s_{ij}}$  = factor score ( $i$ ) of sample ( $j$ )

$a_i$  = regression coefficient for factor ( $i$ )

According to Chalmers et al (1977) the use of principal components in regression analysis has the following advantages:

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

- (1) Each regression coefficient can be calculated independently of the others.
- (2) Each component can be tested independently of the others.
- (3) Matrix inversion is simpler.

Since the factors are mutually orthogonal then:

$$a_i = \frac{\sum_j G_j f_{s_{ij}}}{\sum_j f_{s_{ij}}^2}$$

where the summation is over all values of  $j$ .

This process has been applied to the High Street peak data collected on 19/7/72 using the output of rotated factor scores as the independent variables and normalised journey-time as the dependent variable. The correlation coefficient between the journey-time and factor score 1 was 0.700 and with factor score 2 was -.063. For factor 1 this is a highly significant result and the corresponding regression equation is;

$$G = .700fs$$

Substituting the original normalised variables into this equation and using the factor score coefficients calculated previously gives;

$$fs = .445(Z1) + .294(Z2) + .406(Z3) + .131(Z4)$$

where  $Z1$  = normalised variable  $X1$

$X1$  = traffic flow

$X2$  = crossing pedestrians

$X3$  = turning vehicles

$X4$  = opposing traffic flow.

After substituting the appropriate values of  $X1$ , the resulting equation forecasting  $Y$  in terms of  $X1$ ,  $X2$ ,  $X3$  and  $X4$  is given by;



## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

$$Y = .111(X_1) + .042(X_2) + .017(X_3) + .032(X_4) - .158 \quad (a)$$

The corresponding equation obtained by the usual method of multiple regression analysis, without eliminating non-significant variables, is given by;

$$Y = .174(X_1) - .040(X_2) + .169(X_3) - .072(X_4) + .500 \quad (b)$$

Note that variables  $X_2$  and  $X_3$  do not make a significant contribution to the correlation and would normally be left out of the equation. Forward stepwise regression gives the following equations;

$$Y = .230(X_1) - .500 \quad (c)$$

$$Y = .143(X_1) + .143(X_3) - .435 \quad (d)$$

The values of  $F$  for these equations, in decreasing order of significance, are - (a) ( $F = 15.3$ ), (c) ( $F = 12.7$ ), (d) ( $F = 8.2$ ) and (b) ( $F = 3.9$ ). Thus, on this evidence the model derived by factor analysis gives the closest estimates of journey-time. The essential difference between equation (a) and (b) or (d) is that one regression stage has been carried out for (a), since it is the composite factor score which has been regressed against normalised journey-time, whereas the individual variables have been regressed against journey-time in (b) and (d). In these last two cases the number of stages is equal to the number of independent variables included in the analysis. Another effect of stepwise linear regression analysis is revealed by comparing equations (c) and (d), where it can be seen that the regression coefficient of  $X_1$  changes considerably between the two equations. This occurred because variables  $X_1$  and  $X_3$  were almost equally correlated with journey-time (correlation coefficients of .665 and .653 respectively) and with each other (correlation coefficient = .668). Thus equation (d) provides a non-significant improvement in the accuracy of the estimates over equation (c) and, in fact, a regression equation containing  $X_3$  alone as the independent variable

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

would have a similar level of significance to (c).

An interesting result in equation (a) is that the weights of the variables are all positive. This tends to confirm the hypothesis that an increase in the value of any of these variables would cause an increase in journey-time. This could be a chance result but if the data is typical of this particular section, and other data collected at a peak period confirms this for flow and journey-time, then the pattern would be expected to be similar in other samples. Subjective experiences of the obvious correlations between high pedestrian flows, traffic flows and turning movements also support these observations.

We shall now consider other results which have been analysed using factor models, but first some comments need to be made about the data. The earliest results of the studies of the variation of journey-times in the peak hour showed that traffic flow gave the most satisfactory predictions. Hence later studies were not planned with the intention of measuring the other variables - turning vehicles and crossing pedestrians. Furthermore, the additional data collection would have required a greater number of observers than were available, since the section of Southbury Road which was studied included twelve T-junctions and one zebra-crossing. Thus, further evidence of the existence of peak hour factors is incomplete. More recently further information on the turning flows has been obtained from a cine-film which was taken at the eastern end of the section in the evening peak period on 19/10/73. Two of the junctions on the south side and one on the north side could be seen clearly enough to permit reliable measurements of turning flows to be obtained. No junctions were visible on a film taken at the western end and no information could be obtained of the crossing pedestrians. Because of these limitations on

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

the amount of available data one of the variables - crossing pedestrians - cannot be included in a peak hour factor model. Another variable - turning vehicles - is unlikely to be representative for the whole section because it was observed that about 90% of the turning movements at these three junctions included a vehicle moving into or out of the westbound stream. Bearing in mind these limitations on the accuracy of the data we shall now consider the results. Excluding crossing pedestrians, the factor model derived earlier in this section for the High Street data can be written as follows:

$$fs = .445(Z1) + .406(Z3) + .131(Z4)$$

A linear regression model was derived using the factor scores as the independent variable and the normalised journey-time as the dependent variable. The correlation coefficient between the observed journey-times and the journey-times predicted by the model was .66 for the westbound flow and this result is significant at the 5% level for this sample (sample size - 15). The correlation coefficient of the model using traffic flow as the independent variable was .652. For the eastbound flow the correlation coefficient was only .215 between the estimated and observed journey-times. For this stream no significant relationship was obtained between traffic flow and journey-time using conventional regression analysis.

Another model was derived for this road which was based on factor analysis of three measured variables - traffic flows for the two opposite directions and turning vehicles. The correlation coefficients between the observed and estimated journey-times using this model were .796 and .229 for the westbound and eastbound flows respectively. Considering the westbound flow alone, the results tend to confirm the conclusions derived from the High Street data, but

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

because few variables were measured the models do not provide significantly more accurate estimates than conventional regression models. For the eastbound flow no satisfactory predictive models have been obtained using either conventional regression analysis (see Chapter 3) or the factor method. The results indicate that further studies are necessary in order to determine which variables are most relevant to journey-time models in the peak period, with careful attention paid to the choice of location and the choice of the units of measurement. The latter point is most important since any models produced must be generally applicable. To explain this point further, turning vehicles can be expressed in units of (a) numbers, (b) numbers per junction or (c) numbers per mile. Models which include turning vehicles using (b) or (c) will be generally applicable but not models using (a).

The results of this investigation have shown that factor analysis may be of value to the traffic engineer when studying problems which include several variables. Invariably multiple regression analysis alone has been used in the past and, although investigators may have felt that most variables had some effect, the method of analysis often rejected them or sometimes gave rise to models having the unexpected sign for a particular variable. Factor analysis may permit more variables to be retained to provide intuitively satisfying models without resorting to the use of subjective indices. These models will be more accurate since, as Harris (1975) argues, the factors are based upon stable relationships between the "independent" variables. It is entirely credible that the evidence of the inter-correlations between the "independent" variables is reliable, that is, no less confidence can be placed in these results than those derived by the usual methods

## MULTIVARIATE ANALYSIS OF SPEED-FLOW DATA

of regression analysis. There is no reason to believe that these underlying relationships are less stable than the more obvious ones which have been investigated in previous chapters.

## CHAPTER 6

### OFF-PEAK JOURNEY-TIME STUDIES

This chapter considers a number of studies of off-peak journey-time along various types of road and features of each road which are believed to influence journey-time. The investigations were carried out in the London Boroughs of Enfield, Haringey and Barnet during the period from the summer of 1978 to the spring of 1980. The off-peak "journey-time" used in this study is the total journey-time less any delay at major intersections.

#### 6.1 SITES, METHODS AND PRELIMINARY ANALYSIS.

##### 6.1.1 Sites.

The surveys were carried out along 22.5 miles of single carriageway roads and 11 miles of dual carriageways in Enfield, 8 miles of single carriageways in Haringey and 20 miles of single carriageways in Barnet. Initially large scale plans were obtained of each Borough, and on these the major distributor roads were identified from local knowledge and divided into half-mile sections. The roads are in the Department of Transport's categories of Trunk and Principal roads and so they carry substantial proportions of through traffic. Routes were chosen along roads which passed through different types of areas and varied in length between 8 and 20 miles. The variables which were

## OFF-PEAK JOURNEY-TIME STUDIES

expected to influence journey-time were adjacent land use, road width, number of pedestrian crossings, number of priority-controlled intersections, number of bus-stops, frequency of bus services, traffic flows and number of parked vehicles. Apart from road widths, variations of highway geometry were not included in the study. The reasons are that a close inspection of a large-scale map of the area showed that a large proportion of the roads were straight and, since the journeys were carried out in both directions, it was expected that the effects of gradients would not be evident in the journey-times. In any case, many of the sections of road are quite flat. The measurement of delays at major intersections will be described later. Initially land use was divided into five categories - residential, shopping and commercial, industry, education and other.

### 6.1.2 Methoda.

Mean journey-times were measured during the period from 9-30 to 11-30 am using a variant of the moving-observer method - see Wardrop (1952). The driver was instructed to attempt to drive at a speed such that as many vehicles overtook the observer's vehicle as were overtaken by it. On narrow roads speeds were generally limited by preceding vehicles and on empty roads the driver was instructed to drive at the speed limit and to obey the usual rules at pedestrian crossings and intersections. Journey-times were originally measured using a stopwatch but later a digital stop-clock mounted in the vehicle was substituted since this was easier to use and the possibility of reading errors was reduced. A minimum of five journeys in both directions were carried out each on a separate day. The journeys were

## OFF-PEAK JOURNEY-TIME STUDIES

restricted to Tuesday, Wednesday and Thursday mornings. Since the journeys and routes were continuous the half-mile points were defined by a Halda mileometer reading to 1/100th of a mile and this was set to zero at the start of each run. Where excessive delays occurred the stopped time was measured. In order to estimate journey-times it was necessary to determine the average delays at traffic signals and this was done both by measuring individual delays during each run and also by using the Road Research Laboratory formula (see Road Research Laboratory (1963)). The data for these calculations were obtained by measuring ten consecutive cycle-times, green-times, the flow across the stopline in each cycle and the road width at the stopline. Where long delays occurred at traffic signals the measured delays were preferred to the theoretical delays in the calculations. A number of sections were excluded from the analysis because of excessive delays at traffic signals. These were rejected if, at a particular intersection, the vehicle was delayed for more than one complete cycle on more than half of the journeys. Thus the conditions in these rejected sections were more comparable to those which existed in peak periods and, therefore, the results were not strictly applicable to the conditions appropriate to this study. This is not a completely satisfactory criterion for rejection since the rejected sections tended to be those in which a large number of parked vehicles and zebra crossings occurred and these were usually in busy shopping areas. By rejecting from the analysis those sections which contained high values of some of the most important variables, it is possible that the range of application of the models was reduced. However, such extreme values are unlikely to be typical of such roads in



## OFF-PEAK JOURNEY-TIME STUDIES

off-peak periods and hence their rejection was considered justified.

The road width of each section was determined by averaging the widths at three evenly spaced points. The number of side-turnings was determined by inspection of a large-scale map. Petrol stations, bus-stops and pedestrian crossings were located by marking their positions on the map during a preliminary journey along the routes. Off-peak bus frequencies were obtained from the advertised timetables at the bus-stops.

Two other variables were measured, off-peak traffic flow and the numbers of parked vehicles. Traffic flows were measured by carrying out a half-hour manual count at the mid-point of each section. To minimise the possibility of bias the order of these surveys was randomly determined, and they were carried out during the same three days of the week as the journey-time surveys, but in different weeks because of the large number of surveys which were required. Parked vehicles were counted by an extra observer in the vehicle on each journey and all stopped vehicles on both sides of the road were included, except buses at bus-stops. On some of the wide roads in Barnet where extensive parking is permitted this was sometimes difficult to carry out accurately.

### 6.1.3 Preliminary Analysis

The data were tabulated for each section and discrepancies such as unlikely journey-times were identified and corrected or eliminated. In the early surveys these occurred because of the difficulty of reading the small minute hand on a stopwatch in a moving vehicle. The

## OFF-PEAK JOURNEY-TIME STUDIES

mean journey-times were calculated separately for each direction in order to check for large discrepancies. At this stage a number of sections were eliminated because the effects of oversaturated intersections created great variability in the measurements of journey-time. All of the sections of the North Circular Road which were studied were affected in this way with half-mile journey-times as high as 11 minutes and these results were rejected. As a result the dual-carriageway sample was reduced considerably. After checking, the values were combined and all of the data was punched on to cards for entry into the computer. In the case of the Enfield study, which was the first to be completed, a separate set of data was also collected in sections which overlapped, but were not identical with, the original study. This data was intended to be used to check the validity of any models produced on new data in the old sections before proceeding to test the models in new areas.

The values and ranges of the various parameters are listed in Tables 6.1, 6.2 and 6.3. It can be seen that to a large extent the ranges of the variables overlap between areas although a detailed look at the raw data for individual sections reveals some extreme combinations of values. There are significant differences in the mean values of Bus Frequency and the Number of Side Turnings between the sections in Barnet and Enfield. Between Haringey and Barnet there are significant differences only in the Number of Side Turnings and between Enfield and Haringey there are no significant differences in the independent variables. These comparisons were made in order to ensure that models derived from data from one area could be compared with those derived from the others.

## OFF-PEAK JOURNEY-TIME STUDIES

Preliminary analysis of the Enfield data revealed that because some of the variables have numerically relevant values in only a small number of sections they would be unlikely to provide meaningful models. These tended to be land-use parameters and consequently a number of these were amalgamated. Heavy Industry, occurring in only 4 sections, and Education (3 sections) were amalgamated with the category Other. The variable Pelican Crossings (9 sections) was added to Zebra Crossings to form a combined variable - Pedestrian Crossings. In addition, since vehicles entering and leaving petrol stations were expected to behave in a similar manner to those using minor junctions, these two variables were added to derive the new variable - Side-Turnings. The land-use category "Other" was excluded from any further analysis since the sum of the land-use variables was 100% for each section.

### 6.2 ANALYSIS

#### 6.2.1 Linear Regression Models.

##### Dual Carriageways.

The amount of data for dual carriageways is restricted because there are only about 11 miles of such roads in Enfield, of which a large proportion are subject to excessive delays at a number of busy intersections on the North Circular Road. The results of multiple linear regression analysis are not significant, although they give a weak indication of dependence of journey-time on the off-peak bus frequency. Table 6.4 gives the details. Since these results were obtained early in the studies it was decided to omit dual-carriageways

## OFF-PEAK JOURNEY-TIME STUDIES

from any further analysis.

### Single Carriageways

Because the data was collected over a period of two years the sequence of analysis has been influenced by the order in which the data was obtained. This is the order in which it will be discussed. Firstly, the data collected in Enfield in 1978 was used to determine a number of linear regression models which were tested on further data collected in Enfield and Haringey in 1979. Later, additional data was collected in Barnet during the winter period of 1979-80.

### Enfield Results.

The coefficients of a number of statistically significant models derived from the 1978 data are given in Table 6.5. The models show that both journey-time and journey-speed are highly correlated with the number of pedestrian crossings and the correlation coefficients are .736 and  $-.686$  respectively. On the average each pedestrian crossing increases the journey-time by .18 minutes and decreases the journey-speed by 3.1 mph over a half-mile section. Models containing several independent variables have also been obtained using a forward inclusion process with a test F value of 4.0 to approximate to the 5% level of significance. Models (a) and (b) have been checked against journey-time data collected on other days on 20 sections of the same roads which were used to derive the models. The correlation coefficients between the observed and predicted values of journey-time were .635 and .705 for (a) and (b) respectively. These correlation coefficients are close to the values for the original models, (see Table 6.5).

## OFF-PEAK JOURNEY-TIME STUDIES

A preliminary analysis of the distribution of the number of parked vehicles showed that this was highly peaked and positively skewed (kurtosis 8.27 and skewness 2.69). In order to reduce the skewness this variable was transformed to the square root of the number of parked vehicles; however, neither the original variable nor the transformed variable occurred in any of the significant models produced. This may indicate that parked vehicles do not significantly affect journey-times in Enfield, but another explanation is that parking is correlated with other independent variables (for example, correlation with bus-stops XB is .473) which are more highly correlated with journey-time. Additionally, in most shopping areas in Enfield kerbside parking is severely restricted and, therefore, the effects on journey-time will be difficult to detect. Hence it is not surprising that any independent effect of parking cannot be observed in these results.

Additional data were collected on the same roads in the spring of 1979 in order to carry out an additional check on the validity of the models. The mean journey-time was 1.28 minutes per half-mile section in 1978 and 1.22 minutes per half-mile in 1979. The most significant models derived from the 1979 data are given in Table 6.6. The models marked with an asterisk may be compared with the corresponding models in Table 6.5. In all cases the regression coefficients can be seen to compare very closely although the correlation coefficients are lower than those derived from the 1978 results. No models containing more than two independent variables gave significant correlation coefficients for the 1979 data.

## OFF-PEAK JOURNEY-TIME STUDIES

Two models produced by the 1978 data were used to predict journey-times and speeds; these predictions were compared with the observed values in the 1979 data and the results are given in Table 6.7. The results of the use of equation (c) are also shown in the scatter diagram in Fig 6.1. The models used were those derived from the 1978 data which had high correlation coefficients and were significant at the 5% level on the 1979 data. It can be seen that the correlation coefficients compare very closely with the highest values in Table 6.5. These results indicate that the independent variables used provide stable models within Enfield, and suggest that the models and methods might be more widely applicable.

### Haringey

Multiple linear regression analysis was carried out on the Haringey data and this produced the models shown in Table 6.8. These are the most significant models produced by the data and include variables which were not significant in the Enfield models. Other less significant models were obtained which tend to confirm some of the models produced by the Enfield data and these are compared in Table 6.9. The objective of deriving these less significant models was to attempt to determine if the same variables were likely to be important in the two areas considered separately. These results indicate similarities in the variables which influence journey-time significantly. The 1978 models were also used to predict journey-times and journey-speeds in Haringey and the results are given in Table 6.10. As expected the correlation coefficients indicate agreement which is significant at the 5% level. However, a certain

## OFF-PEAK JOURNEY-TIME STUDIES

amount of caution is necessary in accepting these results since two of the independent variables - Bus Frequency and Bus Stops - have values which extend beyond the range of those in Enfield. As a result one of the significant Enfield models, model (d) in Table 6.5, did not give accurate predictions of journey-times with the Haringey data. This could have been expected since the sign of the regression coefficient of X3 (Bus Flow) does not correspond to the expected sign. The effect of this variable is minimal in the Enfield results but much greater in the smaller Haringey sample. The results of the test using model (c) are shown in the scatter diagram in Fig 6.2.

### Barnet.

Multiple regression analysis of the Barnet data has produced the models listed in Table 6.11. The most significant models include variables which were not significant in the Enfield and Haringey models; these were Shopping Land-use (X5), Parked Vehicles (X13) and Road Width (X2). However, these variables are highly intercorrelated (the correlation matrix being given in Table 6.12) and consequently the regression coefficients vary considerably between the models. Since the individual correlation coefficients between parked vehicles and journey-time, and shopping land-use and journey-time are almost identical, .691 and .696 respectively, the relative significance of these variables in the regression models could change from sample to sample. Provided the samples have been obtained with care models containing either of these variables would be expected to provide results of similar levels of accuracy. But the appearance and disappearance of particular variables can be disturbing. Thus it is

## OFF-PEAK JOURNEY-TIME STUDIES

possible for models using different variables to be produced from different samples, although a close inspection of the correlation matrices might reveal similarities in the correlations between the same variables in different samples.

Other less significant models have also been produced which contain the variables which were shown to be significant in the Enfield models - residential land-use, pedestrian crossings and bus-stops. The intention was to determine the extent to which these variables might be significant in Barnet. The results are shown in Table 6.13 and indicate that a number of models can be derived from this data which have levels of significance which are not much less than those of the most significant models. This highlights a major problem of this type of empirical study, that if several models of similar levels of significance can be derived from the data then the model which gives the highest multiple correlation coefficient for any one sample is unlikely to do this for the others. However, it has been demonstrated in these results that similar models may be applicable to all of the areas although not necessarily the most highly correlated for all samples.

Before continuing with a discussion of models derived from pooled data another potential problem will be considered. Where a number of variables have been measured in more than one population, or separate groups of samples, it is important to determine whether there are significant differences in the values and ranges of the variables between the separate populations. The reasons are twofold. Firstly, if significant differences occur then in theory a linear regression



## OFF-PEAK JOURNEY-TIME STUDIES

equation derived from one population may not be applicable to the others. In practice the researcher would need to exercise some judgement on the application of such a model depending upon the extent of the differences and the importance of the variables to the model. Secondly, where alternative forms of model are available such an analysis can give an indication of the variables which should be used in the model in order to make it more widely applicable. The process of determining the significance of the differences between the mean values of two or more multivariate populations is known as discriminant analysis.

Where only a single variable has been measured a discriminant analysis can be carried out using the t test. However, when several variables have been measured more complex methods need to be used since it is the coincidental variation within a number of dimensions which needs to be assessed. We can perceive such a method geometrically as determining the locations of the mean values of each population within the hypothetical space defined by the variables. The populations will be considered similar if the distance between the mean values is less than some proportion of the common variance. Thus this kind of analysis is similar to the t test. The methods used are capable of allowing for inclusion or exclusion of variables in the discrimination process depending upon the extent of the improvement in discrimination which results. For further reading on discriminant analysis see Bennett et al (1976) or Cooley et al (1971).

A discriminant analysis was carried out between each pair of samples of data and also between the three samples considered together.

## OFF-PEAK JOURNEY-TIME STUDIES

Firstly, an analysis of variance between the groups was carried out on the individual variables, the data for each Borough being treated as a separate group. The next stage required the calculation of the linear relationships of the variables which were needed to separate the groups and which were most significant in identifying an individual sample with the correct group. With pairs of groups one linear function is required to discriminate between them and with three groups two functions are needed. With pairs of groups the function corresponds to the distance parallel to the line passing through points defined by the mean values of the variables for each group in many dimensions. The number of dimensions corresponds to the number of variables included in the function. The results show that no discriminant function can be devised to discriminate between the data collected in Haringey and that collected in Enfield with a reasonable level of accuracy. But between Barnet and Enfield the number of side-turnings, road-width and bus-flow varied significantly, as they did for the three groups considered coincidentally and these variables are contained in the discriminant functions. These functions identified 59 samples out of 99 with the correct group. However, one of the functions had a low level of significance and this result suggested that the data for Enfield and Haringey could be combined when the result of the earlier analysis between these two was taken into account. After combining these two a further discriminant analysis indicated that the only variable which was significantly different between the Barnet data and the combined Enfield-Haringey data was the number of side-turnings. An interesting result of these discriminant analyses is that the variables which are most significant

## OFF-PEAK JOURNEY-TIME STUDIES

in discriminating between the samples rarely occur in the regression equations. This suggests that models could be derived from the pooled data which would not be rejected if tested on one of the samples separately, and which are more likely to be widely applicable than any of the models derived from an individual sample. However, it must be noted that the data for each Borough is not a random sample from the pooled data, hence any predictions produced by the pooled models on the data for one Borough only will not be as accurate as the "best" models derived earlier for the individual areas.

### Pooled Data.

Multiple regression analysis of the pooled data produced the models which are listed in Table 6.14. Because a number of variables have been shown to be highly correlated with journey-time and these have tended to be the same variables in all three areas, these variables also occur in the most significant models in the pooled data. Thus the variable which is most highly correlated with journey-time in the pooled data is the number of pedestrian crossings, with a correlation coefficient of .619. This can be compared with correlation coefficients of .541, .736 and .449 in the Barnet, Enfield and Haringey samples respectively. In the models containing two independent variables the number of pedestrian crossings can be paired with either the residential land-use or shopping land-use, with only a small change in the multiple correlation coefficients, .667 and .694 respectively. It is unlikely that satisfactory regression models containing both residential land-use and shopping land-use can be devised since these two variables are highly negatively intercorrelated (correlation coefficient:  $-.657$ ).

## OFF-PEAK JOURNEY-TIME STUDIES

The scatter diagram for the most significant model containing three variables is shown in Fig 6.3, the multiple correlation coefficient being .716. This model has been tested separately on the three samples of data, giving correlation coefficients between the observed and predicted journey-times of .721, .733 and .719 on the Barnet, Enfield and Haringey data respectively. These results have shown that there are significant, consistent relationships between journey-time and the following variables - numbers of pedestrian crossings, shopping land-use and numbers of bus stops. The consistency of these results suggests that it is probable that highly significant and more widely applicable results could be obtained from a more extended study, even if the areas were different in character from each other.

### 6.3 EFFECTS OF SPATIAL CORRELATION

At the beginning of these off-peak studies a decision was made to fix the length of each survey section at a half-mile. The choice of that length was arbitrary, since it is not immediately evident which is the most suitable length for study. In other studies, for example see Wardrop (1952) and Freeman, Fox et al (1972), no attempt was made to choose spatial sampling intervals of equal length, presumably because the researchers did not expect that the differences in journey distances would seriously affect the results. For this study the possibility of such effects occurring was considered, hence the decision to make all the space sampling intervals of equal length. The extent to which such effects are evident in the results will now be considered.

## OFF-PEAK JOURNEY-TIME STUDIES

In Chapter 3 the consistent effects of aggregation of journey-time/flow data into successively longer sampling intervals was discussed with respect to data collected in consecutive intervals of time. Wright (1971) suggests that such effects might also be observed in data which has been measured in space sampling intervals, and in this section these effects will be considered. In order to permit the regression coefficients to be compared directly each variable was expressed as the average per half-mile. The aggregation is similar to that carried out for the peak-hour studies, but because the sampling domains are different a number of other differences arise. Firstly, in the peak hour the mean journey-times are weighted according to the traffic flows in the individual sampling intervals but in the off-peak data no such weighting occurs. This implies that Bias Type 1 will be absent from the off-peak results. Secondly, in the peak hour Bias Type 2 arises because of time-lagged cross-correlations between journey-time and flow. If space-lagged cross-correlations occurred Bias Type 2 would be expected in the off-peak data.

The results of this analysis are given in Table 6.15, and are difficult to interpret. The regular trends which were evident in the peak-hour results are no longer clearly defined, although the coefficients for the longer sampling intervals are larger than those in the shorter intervals. This result is only partly consistent with the predictions of Wright's models discussed earlier.

Other indirect evidence for the existence of spatial correlation has been obtained by analysing the results of the study reported by Bampfylde (1979). He analysed the variations of journey-speed and

## OFF-PEAK JOURNEY-TIME STUDIES

traffic flow in a number of tunnels of various lengths in Britain. He obtained the wide scatter of speed/flow results which many other observers have obtained, and was unable to establish relationships which were uniquely applicable to tunnels. However, the tunnels were of widely different lengths, from 183 metres to 3138 metres, and there were differences in the vertical and horizontal curvature, but these differences had no detectable effects on journey-times. If the regression coefficients of speed on flow given by Bampfylde are compared with the tunnel lengths it will be found that they are correlated. The correlation coefficient is  $-0.575$  and the F value is 5.6. Furthermore, it is most likely that each tunnel was of a constant width throughout, hence the characteristics were probably similar along the full length of each tunnel apart from vertical and horizontal curvature.. Thus if each tunnel had been divided into a number of equal sections the characteristics of each section would have been almost identical with adjacent sections and Bias Type 2 would be evident in the results. This is not a convincing argument since it is possible that the regression slopes of the relationships between traffic flow and journey-speed might be expected to become steeper with increasing length of tunnel simply because of the effect of side friction, and this would be expected to be greater in a long tunnel. Generalising Wright's hypothesis, if the journey-time of the traffic in a section (sampling interval), is correlated with some independent variable occurring in some nearby section (lagged interval), then the slope regression coefficient between journey-time and that variable is likely to change in a consistent manner as the variables are aggregated into successively longer sampling intervals.

## OFF-PEAK JOURNEY-TIME STUDIES

Clearly the hypothesis is appealing although the evidence presented here is not conclusive since relatively small samples have been used; however, the results are statistically significant at the 5% level. The results suggest that care should be taken when comparing two studies to ensure that the space sampling intervals are similar in size, otherwise the comparisons may be invalid.

## CHAPTER 7

### MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

#### 7.1 FACTOR ANALYSIS MODELS

This part of the study of off-peak journey-times was planned with the objective of creating and testing a number of indices of land-uses and other variables alongside the road. That is, it was intended to find out if the characteristics of the features along the road could be expressed quantitatively using an index based on measurements of particular variables and the inter-relationships between them. Furthermore, it was intended to study the relationships between journey-time and such an index. One such index has been postulated by Buchanan et al (1976) in a study of journey-times on roads in London in 1971. This was called an Activity Index and was expressed by:

$$AI = 0.10 \text{ SHOPS} + 0.02 \text{ HOUSES} + 0.01 \text{ NBU} + 0.05 \text{ OBU}$$

where SHOPS = % of both sides of road occupied by shops,

HOUSES = % of both sides of road occupied by houses,

NBU = % of both sides of road not built up,

OBU = % of both sides of road occupied by other premises.

Buchanan's paper gives a reference to Kinloch (1972) as the source of the Activity Index; however, the latter paper merely quotes the formula and includes a graph showing how the Activity Index varies with certain



## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

parameters. Further searches have not revealed the justification for the particular weightings given to the individual variables in the expression although the sampling methods used for measuring the variables are given by Large (1971). The calculations of Activity Index in the reference seem to indicate that the weightings are arbitrary and based upon some subjective perception of the influence of land-use activities on journey-times. In the present study the methods of principal component and factor analysis have been used for the purpose of investigating relationships between land-use variables and for deriving indices. These methods are described in Appendix E.

As was stated in Chapter 5, in the method of principal component analysis the first component is so derived that it accounts for the maximum amount of variance in the observed data. Further components are successively derived in such a way that they account for the maximum amount of residual variance and they are orthogonal to any components already derived. The proportion of variance accounted for by each component is expressed by its eigenvalue. It can be shown that since the variables are normalised at the start of the process then the sum of the eigenvalues will be equal to the number of variables. Kaiser (1958) argues that the important components will be those whose eigenvalues exceed 1.0, although this is by no means a strict rule. In this study nine variables have been measured and these are; road width, frequency of off-peak buses, residential land-use, shopping land-use, numbers of bus stops, numbers of side turnings, off-peak traffic flow, numbers of parked vehicles and numbers of pedestrian crossings. Some important features of the relationships between the components and the raw variables are shown in a factor pattern matrix. This is a matrix of

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

correlation coefficients, called factor loadings, between the variables and the components. Like the more commonly used Pearson correlation coefficients high values of the factor loadings indicate significant relationships between the variables and the components. Furthermore robust relationships should give similar factor patterns with different samples from the same population. Therefore, in these studies identical types of analysis have been used on each of the separate samples and on the pooled data.

Principal component analysis of the Barnet data shows that the three most important components have eigenvalues of 4.63, 1.35 and .93 respectively (see Footnote). These three components account for about 77% of the total variance of the nine variables and for proportions of the variance of individual variables of between 61% for pedestrian crossings, and 83% for road width. The first principal component has high loadings (correlation coefficients) on all of the variables, which is usually the case, and is therefore difficult to interpret. Consequently these three principal components were then subjected to varimax rotation, and the results enabled individual factors (rotated components) to be identified with particular sets of variables, (see Chapter 5). The results of this rotation provide evidence for the existence of three interpretable factors. Thus Factor 1 was found to be positively correlated with shopping land-use, pedestrian crossings and Footnote. To avoid unnecessary repetition of similar tables sample output is not given for the three areas considered separately but only for the pooled data, and these results are shown in Tables 7.1 and 7.2. The results are discussed later in this chapter.

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

parked vehicles, and negatively correlated with residential land-use. The inclusion of these variables in this factor is not surprising since it might be expected that there would be more frequent pedestrian crossings, heavier car-parking and less residential land-use in shopping areas. The rotated factor matrix succeeds in quantifying this association between the variables. Factor 2 is most heavily loaded on road width, bus frequency and traffic flow and moderately loaded on parked vehicles. Factor 3 is loaded on bus frequency and side-turnings.

For the Enfield data the first three principal components have eigenvalues of 3.14, 1.97 and 1.17 respectively and they account for about 70% of the total variance in the data. These three components account for a minimum of 55% of the variance of parked vehicles up to a maximum of 78% of the variance of shopping land-use. Factor 1 was highly correlated with shopping land-use, traffic flow and pedestrian crossings, and negatively correlated with residential land-use. Factor 2 was heavily loaded on side-turnings and bus-stops. Factor 3 was loaded on road width, bus frequency and parked vehicles. Parking restrictions are more severe in shopping areas in Enfield than in shopping areas in Barnet, hence the lower correlation with the other variables associated with factor 1. The variables which are most highly correlated with the other two factors are remarkably similar in both sets of data although the relative importance of these two factors changes between the samples. The results of this analysis of the Enfield data tend to confirm the conclusions derived from the Barnet data.

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

In the analysis of the Haringey data the first three principal components had eigenvalues of 3.55, 1.55 and 1.16 and these accounted for about 70% of the total variance. The first factor was highly correlated with bus frequency, shopping land-use and bus-stops, and was negatively correlated with residential land-use. Factor 2 was highly correlated with side-turnings, number of pedestrian crossings and road width. Factor 3 was highly correlated with car-parking. Thus these results differ from the previous ones although two of the variables associated with factor 1, shopping land-use and residential land-use, also occur in the corresponding factor in the Barnet and Enfield data. It is possible that in such a small sample one or two extreme values may have given rise to the differences in the structure of the factors between these results and those derived from the Barnet and Enfield data.

The analysis was then carried out on the pooled data and, as would be expected, gave results which were similar to those obtained by the analysis of the Barnet and Enfield samples. The first three principal components had eigenvalues of 3.59, 1.51 and 0.90 respectively. These three components accounted for 67% of the variance of the data. The first factor was highly correlated with shopping land-use and pedestrian crossings, and negatively correlated with residential land-use. These results are shown in Table 7.1. Factor 2 was highly correlated with road width, bus flow, bus-stops and traffic flow. Factor 3 was highly correlated with side-turnings and parked vehicles. Thus the results suggest the existence of a distinct bipolar factor positively related to shopping land-use and frequent pedestrian crossings on the one hand and negatively related to residential land-use on the other; this

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

corresponds to factor 1 in this analysis. The second factor is associated with bus frequency, numbers of bus stops and road width. Factor 3, because it is mainly associated with a single variable "side turnings," can be regarded as a unique factor.

The importance of each factor is indicated by the corresponding eigenvalue, which gives the amount of variance in the data explained by that factor. The sum of the eigenvalues is equal to the number of variables in the data set. For the discussion above the three most important factors were considered, but one of these has an eigenvalue less than 1.0 and most analysts would not regard that factor as worthy of further analysis. If that factor is omitted then the factor pattern can change quite considerably, since fewer factors are being used to explain the variance of the full set of variables. In chapter 5 it was stated that the proportion of the variance of an individual variable, which can be accounted for by the variation of the common factors, is known as the communality of the variable. In general, the communalities of all of the variables will be reduced if a factor is omitted from the analysis, but the effect on an individual variable will depend upon its correlation with the omitted factor. The effect on the results of this study of using two factors instead of three is to make the two remaining factors account for a greater proportion of the variance after rotation than the first two rotated factors of the three-factor model. The greatest change occurred in factor 2; in the two-factor model this factor accounted for about 50% and 45% of the variance of side-turnings and parked vehicles respectively but only 6% and 5% in the three-factor model. In most other respects the interdependence between the factors and the remaining variables, as shown by the factor loadings, did not

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

change very much and this can be taken to indicate that these two factors are stable.

A convenient method of displaying these results is to use a vector diagram of factor loadings. The diagram showing these results is given in Fig. 7.1. The axes of the diagram represent the first two principal components PC 1 and PC 2. The corresponding positions of the rotated factors are labelled F1 and F2 and the projections of these factors on the original axes represent their corresponding correlations with the two principal components. The positions of the points corresponding to the raw variables are defined by their correlation coefficients (factor loadings) with the principal components. The correlations between the variables and the rotated factors are given by the projections of the variables on the rotated axes. Since the diagram displays correlation coefficients the periphery of the domain is defined by a circle of radius 1.0. The information which the diagram conveys by the close clustering is the grouping of the variables which are highly associated with individual factors. Thus variables X5, XS and X4 are located close to the axis F1; and X2, XB and XT are close to axis F2. Because X4 is negatively correlated with F1, it is shown close to axis F1 but on the opposite side of the axis labelled F2 from the variables X5 and XS, which are both positively correlated with F1.

As we have seen the results indicate that statistically significant relationships exist between the variables and these relationships have been used to postulate the existence of two common factors which cannot otherwise be measured or observed. The first factor is associated with shopping land-use, residential land-use and frequency of pedestrian

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

crossings and the second factor with bus-frequency, numbers of bus stops and road width. Harris (1975) argues that provided the usual precautions are exercised in sample surveys then such factors are likely to be stable and significant. Thus, although it may not be possible to say that variations in the variables are caused by variations in the factors, in the absence of a clearly identifiable theory these factors can be used as indices of the associated variables which they contain. These indices express some underlying relationship between the variables. Finally the factors have been shown to contain the effects of a range of variables which have been used as independent variables in multiple regression models, and this suggests that the factors could also be used in such models.

### 7.2 REGRESSION ANALYSIS OF FACTOR SCORES

Principal component analysis and factor analysis are usually used to explore the inter-relationships within data structures between a wide range of variables. However, some authors, for example Jolliffe (1973) and Kendall (1975), have discussed the use of the factors produced as independent variables in linear regression analysis with other parameters as the dependent variables. This has been done for this study and the independent variables used were the factor scores derived from the following relationship:

$$Fs_{ij} = \sum f_{ik} \cdot z_{jk}$$

where  $Fs_{ij}$  = factor score of the  $i$ th factor in the  $j$ th sample,

$f_{ik}$  = factor score coefficient of the  $i$ th factor  
on the  $k$ th normalised variable  $X_k$ .

$z_{jk}$  = normalised value of variable  $X_k$  in the  $j$ th sample.

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

These factor scores were calculated for each sample in the data bank. The dependent variables were normalised journey-speed and normalised journey-time.

The advantage of using factor scores for regression analysis is that they are mutually orthogonal if derived by principal component analysis, and are nearly orthogonal if derived by factor analysis. Hence the regression equations are easier to interpret since the correlation of the dependent variable with a series of such orthogonal independent variables is the sum of the separate correlation coefficients. In addition, the slope regression and correlation coefficients are identical if normalised data are used.

The correlation matrix for this analysis using a three-factor model is shown in Table 7.2. BJS is the normalised journey-speed and BJT is the normalised journey-time. The table reveals that the factor scores derived from Factor 1 are highly correlated with journey-time and produce a regression model having a level of significance comparable to those models given by regression analysis of the raw variables. The details of the most significant relationships are given in Table 7.3 and the predictions for one of the models are shown in the scatter diagram in Fig. 7.2. The predicted journey-times (PJT) are given by:

$$\text{PJT} = 1.26 + .168F_1 + .052F_2 \dots (A)$$

$$R = .668; \quad F = 38.7; \text{ Standard Error} = .198 \text{ min.}$$

The factors  $F_1$  and  $F_2$  were derived from the first three rotated principal components. The correlation between the observed and



## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

predicted journey-times is highly significant. Three factors have been used in deriving the above model largely for the sake of consistency with the earlier discussion although using Kaiser's guideline for selecting the appropriate number of factors, the number of components whose eigenvalues exceed 1.0, suggests that only two such factors are needed. The corresponding relationship derived from the first two rotated principal components,  $F1'$  and  $F2'$ , is given by:

$$PJT = 1.26 + .173F1' + .047F2' \dots (B)$$

$$R = .678; \quad F = 40.9; \text{ Standard Error} = .196 \text{ min.}$$

As expected these two models display similar regression coefficients because each raw variable has similar factor loadings in both sets of rotated principal components, therefore  $F1$  is similar to  $F1'$  and  $P2$  is similar to  $F2'$ . The regression equations can now be written in terms of the original variables by substituting for each factor the corresponding linear equation which produces the factor score. Thus equation (1) above becomes:

$$\begin{aligned} PJT = 1.22 - .0007X2 + .0019X3 - .0022X4 + .0023X5 \\ - .0074XT + .0461XS - .0050XB + .0001X11 - .0002X12 \end{aligned}$$

And equation (2) becomes:

$$\begin{aligned} PJT = 1.23 - .0004X2 + .0022X3 - .0021X4 + .0022X5 \\ - .0017XT + .0498XS - .0038XB + .0001X11 - .0005X12 \end{aligned}$$

Later in this chapter strategies for reducing the number of raw

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

variables in these models will be discussed.

Since Factor 1 has been shown to be highly correlated with some of the variables which Buchanan et al (1976) used in their Activity Index, it can be seen that the principal component method described above produces a similar kind of index. However, in this case the weightings given to the individual variables are derived in a much less arbitrary manner and represent a known relationship between the variables. An interesting application of the Activity Index was its use as an independent variable in a regression model to predict journey-speeds in London where a high correlation coefficient of .81 was reported. The model has been used in other transportation studies in London and West Yorkshire.

There are two further comments on Buchanan's index. Firstly, the variables defining the Activity Index lead to redundancy - since the sum of the percentages of land-use will be 100% for all sections of roads. Secondly, six-minute sampling intervals were used when measuring the journey-times, and the mean journey-time within each sampling interval was treated as an independent estimate, whereas the sampling intervals for the Activity Index (and the other independent variables used) were lengths of road. The use of each six-minute mean journey-time as a separate independent variable is incorrect because the model was intended to be used for predicting changes of mean journey-time between sections of road. If the overall mean journey-time had been used for each section it is likely that different regression coefficients would have been obtained, and the correlation coefficient would have been based upon fewer degrees of freedom.

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

The current study, using an index derived from factor analysis, avoids the latter two criticisms by the correct choice of variables. However, the choice of weightings is a major weakness of any index unless there is some fundamental underlying hypothesis on which it is based. Since factor analysis seeks to identify the inter-relationships between the variables and, in this study, has found at least one factor which has a plausible, recognisable and possibly causal link with the variables with which it is associated, it seems likely to lead to less subjective indices. Provided the samples have been carefully selected the underlying inter-relationships on which the factors are based are no less likely to be reproducible than any other empirical models. Furthermore the factors derived in this study can be easily checked since the variables do not change very rapidly in time and many of them may remain constant for many years.

The factors which have been used in the regression models discussed above have each contained the effects of all of the variables to some extent. Furthermore, since rotated factors have been used some of the factor loadings were very small, and this is generally true for all rotated factor patterns. Consequently some of the variables have a negligible effect on the factor scores and would not cause much change in the factor patterns if they were omitted. In the remainder of this chapter various strategies for the elimination of potentially unnecessary variables from these models will be discussed.

A number of researchers (Jolliffe (1972),(1973); Dsling and Tamura (1970); Massy (1971); Mansfield et al (1977)) discuss methods of omitting unnecessary variables from factor models used in multiple

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

regression analysis. Jolliffe discusses in detail a number of strategies which can be used for discarding variables and gives the results of tests carried out on both artificial and real data. The results show that for artificial data several rejection methods successfully discarded those variables known to be redundant. When the rejection methods were used on real data none was overwhelmingly superior to the others in selecting sets of variables which provided the most meaningful regression models, when compared with regression models determined by the more common multiple regression procedure.

Two of the strategies described by Jolliffe use the methods of principal components to aid the choice of variables, and these strategies are labelled B1 and B2. Another method calculates the multiple correlation coefficient between each variable and the remaining variables, and uses this as the criterion for selection or rejection; this method he labelled A2. Methods B1 and B2 have been tested on the pooled data in this study. In method B1 a principal component analysis was carried out on the raw variables. The number of components having an eigenvalue greater than 1.0 was noted - in this case, two - and this determined the number of variables to be retained. Then the variable having the highest correlation with the least important component was eliminated. These processes of principal component analysis and variable elimination were repeated several times until there were two remaining variables, residential land-use (X4) and numbers of side-turnings (XT). The two principal components were then rotated using the varimax method. A multiple regression analysis was then carried out using the normalised dependent variables and the factor scores as the independent variables. In method B2 raw variables were eliminated after the first principal

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

component analysis by discarding those variables which had the highest correlation with components having eigenvalues less than 1.0. The remaining variables were residential land-use (X4) and bus frequency (X3). A further principal component analysis and varimax rotation were carried out on the two remaining raw variables, and the remainder of the analysis was identical to B1. The results of these analyses are shown in Table 7.4.

The following observations can be made on the results of these calculations;

(a) the selection methods retained one variable in common, residential land-use; (b) the range of the eigenvalues of the factors using two variables is narrower than that of the factors with nine variables, this being expected since the two variables are the most independent of the original variables; (c) although significant multiple correlation coefficients were obtained these were of lower significance than those obtained by linear regression analysis of the raw variables.

Observation (c) would be expected since the principal component analysis elimination procedures generated models which included fewer variables than, and sometimes different ones from, those which occurred in a number of the most significant linear regression models derived from the raw variables. Furthermore, since the most independent variables were retained the rotated factors tended to be highly associated with one variable. Since factor analysis and principal component analysis seek to find evidence of factors associated with a large number of variables, it seems illogical to discard a large proportion of the variables which

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

provide that evidence.

Thus, the principal problem which has arisen in using these methods of variable elimination is that in each case too many variables have been eliminated. This could have been caused by using a criterion for variable elimination which was too strict or using it at an inappropriate stage. In these methods the number of variables to be retained was determined by the number of principal components whose eigenvalues were greater than 1.0. However, if we determine the eigenvalues of the factors (rotated components) it will be observed that they become more even than they were. For example, the eigenvalues of the first two principal components were 3.59 and 1.51 but the eigenvalues of the factors were 2.59 and 2.51 in the two-factor model. Thus, when a principal component has an eigenvalue which is just less than 1.0, if that component is retained in the model then, after rotation, it is likely that all of the factors will have eigenvalues greater than 1.0. If we now use this increased number of factors to determine which variables should be retained it is clear that fewer variables will be eliminated and, possibly, different variables from those eliminated in the two-factor model. In this data, the third and fourth most important principal components had eigenvalues of .90 and .80 respectively, and so it is likely that after rotation the corresponding factors would have eigenvalues greater than 1.0. Another disturbing effect of these elimination procedures was observed when a variable, which was highly correlated with journey-time, was eliminated because it was also highly correlated with an unimportant component. In this data, the variables X5 (shopping land-use) and X8 (pedestrian crossings) were both eliminated, although both are highly correlated

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

with journey-time. A more appropriate criterion for rejection needs to be found so that the variables retained make a significant contribution to the variation of the factors. And, if the regression models derived by regression analysis of the raw variables and of the factors are to have similar levels of significance, then the factors should retain at least the same number of variables as the regression models derived from the raw data. In general, the retained variables should be those having the highest correlation coefficients (factor loadings) with the factors. Perhaps at this stage we should note that there exists an unresolved controversy on whether factor loadings or factor score coefficients should be used when interpreting factors. Some of the standard texts, see Cooley et al (1971) and Kendall (1975), state that the factor loadings should be used for interpretation of factors. Whereas Harris (1975) gives the counter argument that factor score coefficients should be used for this purpose. These two alternative methods do not necessarily lead to the same conclusions. Furthermore, since some of the variable elimination strategies can sometimes lead to very few variables being retained, the possibility of misinterpretation becomes more likely. Whatever opinion is held on interpreting factors, the variables retained and their number can have a significant effect on the linear regression models derived from the factor scores.

In elimination process type A2 the multiple correlation coefficient of each independent variable with all of the other independent variables is first calculated, giving as many multiple correlation coefficients as there are variables. Then the variable having the highest multiple correlation coefficient is eliminated. Jolliffe argues that the variable which is eliminated at this stage subtracts the least amount of

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

information because of its high correlation with the remaining variables. The process is repeated with successively fewer remaining variables until the multiple correlation coefficients lie below some limiting value appropriate to a selected level of significance. The results are shown in Table 7.5. This method of variable elimination resulted in only two variables being retained - residential land-use and numbers of side-turnings. These are the most mutually independent of the possible pairs of raw variables and their correlation coefficient is only .005. With these two variables, which are virtually independent, factor analysis is irrelevant and multiple regression analysis of the raw variables, with journey-time as the dependent variable, can be used to produce a significant model. But the resulting model only includes one independent variable - residential land-use, and has a correlation coefficient of .45. The inclusion of the variable "side-turnings" does not improve the correlation coefficient significantly. Clearly the number of variables eliminated by method A2 can be altered by choosing a different limiting value for the correlation coefficient. In these results a value of 0.19 was selected; with a sample size of 99 this coefficient is just significantly different from zero at the 5% level. Additional variables could have been retained by choosing a higher level of significance but it is not immediately evident that a higher level is justified. It is clear that this variable elimination procedure is essentially a method of selecting raw variables to be used in regression analysis, not in factor analysis, since the implied independence of the remaining variables means that factor analysis is inappropriate.

From the results of the strategies tested so far no one method of elimination of variables appears to be clearly superior to the others,



## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

or even to the use of stepwise selection methods in multiple regression analysis. This may be a result associated with this particular data, or perhaps more stringent criteria for rejection of variables were used than are justified. But in some of the methods the criteria for rejection or acceptance are arbitrary, and this does not give the user much confidence in the methods. Furthermore, some of the methods always retain only one variable per factor and this can lead to too few variables being retained if only a few significant factors are evident in the data. It seems illogical to use the method of principal components, which seeks to detect the presence of latent factors in the data, and then to eliminate a large proportion of the variables which provide evidence of the factors.

One other published method of variable elimination will now be mentioned, that devised by Mansfield et al (1977). This method combines an attempt to retain variables which maintain the simple structure of the factor patterns (see Appendix E) and provide the most significant correlation with the dependent variable in a regression analysis model. Moreover this is carried out in a single stage. The disadvantage is that the method requires the use of complex matrix manipulation which is very difficult to program. Consequently the present author has devised a two-stage process of variable elimination which tries to achieve the same kind of result using standard computer programs. This process will now be described.

Kaiser (1958) argues that if the factor patterns indicate that the variables are closely clustered then the factors can be defined with fewer variables with little change in their patterns. This he proved

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

analytically for a two-factor model and demonstrated empirically for a model containing several factors. By closely clustered he meant that each factor was highly correlated with a distinct subgroup of variables. Close clustering of the variables is evident in the results of this study and consequently the principle established by Kaiser has been used to eliminate potentially redundant variables in the factor models. This variable elimination process aims to preserve the important characteristics of the factor pattern, and hence the interpretation of the factors, but with fewer variables. A further objective is to obtain a regression model having a high correlation between the dependent variable and the factors included in the model. The important characteristics of the factor pattern which should be retained are the high loadings of the factors on some of the variables and it is these variables which should be retained. The proportion of the variance of a single variable which is explained by the factors is expressed by the communality of that variable. Thus a low communality indicates that a variable has a low correlation with the factors. Hence, in this variable elimination process we will eliminate at each stage that variable having the lowest communality with the rotated factors. Initially the communalities were determined after a principal component analysis and varimax rotation of the first two principal components since these were the only ones having an eigenvalue greater than 1.0. At this stage the variable "traffic flow" was eliminated. A further principal component analysis and varimax rotation were carried out on the remaining variables which led to the elimination of the variable "road width". At each stage the factor scores were generated and these were used as the independent variables in a multiple linear regression

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

analysis with normalised journey-time as the dependent variable. The results of these analyses are given in Tables 7.6 and 7.7. The process was continued until only three variables remained, although it is probably desirable to retain a greater number of variables in order to maintain most of the structure of the factors.

Two alternative criteria can be suggested for choosing the stage at which the variable elimination process is stopped; (a) if the elimination of another variable leads to a significant change in the factor pattern or (b) if there is a significant reduction in the multiple correlation coefficient. The results of the multiple regression analysis of normalised journey-times with the factor scores derived from the rotated principal components are given in Table 7.8. The tables indicate that (a) the factors retain their underlying patterns as the number of variables is reduced; and (b) the multiple correlation coefficients of the regression models do not change significantly. In addition it appears that the multiple correlation coefficients are influenced more by the number of factors used in the models rather than the number of variables used in specifying the factors. Using the coefficients and variables calculated by the program the regression models derived from factors can be written in terms of the original variables. This was carried out for the three-variable factor model and produced the equation given below:

$$PJT = 1.19 + .0035X5 - .0028X4 + .0322XB$$

$$R = .643; \quad F = 33.8; \quad \text{Standard Error} = .204 \text{ min.}$$

Multiple regression analysis of the same three raw variables gave the

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

following model:

$$PJT = 1.11 + .0044X5 - .0017X4 + .0322XB$$

$$R = .651; \quad F = 23.2; \text{ Standard Error} = .192 \text{ min.}$$

In these models it can be seen that similar correlation and regression coefficients were obtained; however, using the normal methods of regression analysis the second model would not have been derived. This is because in stepwise multiple regression analysis, if reasonable values of the test statistic are used, then the process ceases at the stage when XS, X5 and XB have been included in the model. The resulting equation is given by:

$$PJT = .95 + .103XS + .0035X5 + .027XB$$

$$R = .716; \quad F = 33.3; \text{ Standard Error} = .187 \text{ min.}$$

Using the six-variable factor model, since this was the two-factor model which produced the highest correlation coefficient and level of significance, gave the following model:

$$PJT = 1.16 - .0022X4 + .0024X5 + .0625XS \\ + .0049XB + .0042X3 + .0028XT$$

$$R = .688; \quad F = 43.0; \text{ Standard Error} = .193 \text{ min.}$$

The corresponding linear regression model is given by:

$$PJT = 1.05 - .0010X4 + .0029X5 + .1029XS \\ + .0321XB - .0001X3 - .0071XT$$

$$R = .727; \quad F = 17.2; \text{ Standard Error} = .187 \text{ min.}$$

There are greater differences between the coefficients of the variables

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

in these two models than those using fewer raw variables, and furthermore variables X4, X3 and XT would not have been included in the linear regression models if reasonable values of the test statistics had been used. However, from an examination of the standard errors it can be seen that similar levels of accuracy are obtained by both models.

Although the predictions of journey-time using models derived by factor analysis are no more accurate than those derived by conventional linear regression analysis, there are advantages in using such models. Firstly, the numerical values of the factors may be used to compare the characteristics along different roads. Secondly, the use of such factors as independent variables in linear regression models justifies the use of a larger number of raw variables in the predictive models, and these are therefore likely to prove to be more widely useful since the effects of a greater number of variables can be taken into account. Although the methods appear to be complex a number of well-known computer packages of statistics programs include factor analysis subroutines and these are no more difficult to use than multiple linear regression subroutines.

The factors are not directly measurable but are, nevertheless, reasonably consistent expressions of the numerical relationships between the raw variables. These factors can be used where fundamental theoretical relationships do not exist, usually where the variables have been selected for their ease of measurement rather than that they occur in some established hypothesis; this is the case in this study. A disadvantage of using such factors is that there are no rigorous statistical tests for determining the correct number of factors to use

## MULTIVARIATE ANALYSIS OF OFF-PEAK DATA

or the significance of the factor score coefficients, but there are satisfactory tests of significance of the factor loadings. Insofar as inter-relationships exist between the variables factor analysis seeks to find evidence for them in the data and it is implicit in the calculations that strong relationships are evidence of significant latent factors.

## CHAPTER 8

### SUMMARY AND CONCLUSIONS

#### 8.1 Journey-Time/Flow Studies.

Studies were carried out of the relationships between journey-times and classified traffic flows, pedestrian flows and turning vehicles on two roads in Enfield, the High Street and Southbury Road. Separate measurements were made of the flows in the two directions, two sets of data were collected in the off-peak period and six sets during peak periods. The durations of the studies varied between 32 and 90 minutes and the basic sampling interval was one minute. It was found that in the off-peak periods no significant relationships could be detected between journey-time and the other variables in the two sets of data collected in the High Street. Absence of journey-time/flow relationships at low levels of flow have been reported by Wardrop (1952) and Dick (1966).

Unless stated otherwise, in the following discussion the variables traffic flow and journey-time have been measured in a single direction. The separate effects on journey-time of different types of vehicles in the traffic streams could not be isolated because most of the vehicles were cars. In a large proportion of the intervals no buses or motor-cycles were observed; thus in any future studies much

## SUMMARY AND CONCLUSIONS

longer investigations will be needed in order to detect any effect of these vehicles, unless other sections are studied which are known to contain high flows of such vehicles. In the peak periods the results were found to be inconsistent. In the High Street significant linear relationships were obtained between journey-time and unclassified traffic flow in the southbound direction in two of the evening peak periods. With the data aggregated into 5 minute sampling intervals, the correlation coefficients were about .67 and the corresponding slope regression coefficients were about .25 minutes/vehicle/minute. The relationships for Southbury Road were much more varied in their significance and the results depended upon the direction of flow and whether the data was collected in the morning or evening. For the eastbound direction of flow in the evening peak period data there is evidence of the turning-back effect in the relationship between journey-time and flow at high rates of flow, as reported by Rothrock (1957). In the morning peak there is evidence of a significant linear correlation between journey-time and traffic flow. For the westbound direction the morning flows were influenced by a traffic queue which developed at a petrol station and no significant linear relationship was obtained. The turning-back effect was evident in these results and in these circumstances the variations of journey-time cannot be related to the entry flow - see Branston (1976). In the evening a highly significant linear relationship was obtained between journey-time and traffic flow for this direction. Thus this type of study appears to be sensitive to short and longer-term changes in the surrounding conditions and care should be taken in the interpretation of data collected in such cases.



## SUMMARY AND CONCLUSIONS

Different regression coefficients were obtained when the data was analysed in different sizes of sampling intervals in time. In cases where some degree of correlation existed in the data, consistent trends in the coefficients were observed as the data was aggregated into successively longer sampling intervals. The theoretical model derived by Wright (1971) was found to provide reasonable estimates of the regression coefficients for the longer sampling intervals based upon lagged serial and cross covariances in the data collected in one-minute sampling intervals. This observation was confirmed in a number of samples and suggests a means of correcting results for the effects of different sampling intervals when comparing different studies.

Curves having a discontinuity in gradient (called "discontinuous models" in this thesis), non-linear regression models and combinations of these have been tested on each sample. In general, the discontinuous model fitted each sample more closely than a single least-squares line, but the closeness of fit could not be tested by normal tests of significance. A number of problems arose when attempting to fit non-linear discontinuous models. These problems occurred in one of the models because of the need to use the logarithm of the transformed variables which could take negative values, and so this model had to be rejected.

A queueing model which predicts the variation of journey-time with variations of traffic flow was derived by Davidson (1966). In order to account for differences between different roads the model uses observed or estimated values of saturation flow, off-peak journey-time

## SUMMARY AND CONCLUSIONS

and an empirical factor which varies with the surrounding conditions - represented by the symbol  $J$ . This model has been found to give a satisfactory fit to the peak hour data collected in this study. The attraction of this model is that it has a theoretical basis for its derivation whereas many previous models have been entirely empirical. The regular trends in regression coefficients which resulted from using data aggregated into successively longer sampling intervals were also observed in the results derived using Davidson's model. A theoretical argument has been presented which accounts for the trends in saturation flow which were observed in these results. The major advantage of this model is that one equation may be used to predict journey-times over the total range of possible flows. It has been shown that the relationship between journey-speed and a number of variables which was derived by Wardrop (1968) is similar to Davidson's model.

Some authors, see for example Guerin (1958) and Underwood (1960), have discussed the use of models which include concentration as the independent variable in order to predict journey-times or journey speeds. In this thesis attention has been drawn to a fundamental weakness of such models which use concentration derived from flow and speed, or flow and journey-time per unit distance. Although high correlation coefficients are often obtained when the empirical relationships are derived, because these models must be rearranged before using them to predict journey-times or speeds, then much lower correlation coefficients are found between the predicted and observed values.

## SUMMARY AND CONCLUSIONS

The multivariate techniques of principal component analysis and factor analysis have been used to explore inter-relationships in the data collected in the High Street in the evening peak period. The variables studied were those which had been treated as independent for the purposes of multiple linear regression analysis and these are stated at the beginning of this chapter. High correlation coefficients were observed between the "independent" variables traffic flow, turning vehicles and crossing pedestrians. Factor analysis has been used to determine a relationship between these variables and it has been argued that this can be used to define an index of peak-hour activity. There was no evidence of a similar kind of relationship in the off-peak data collected in the High Street.

### 8.2 Off-Peak Journey-Time Studies.

The earlier studies of flows and journey-times in off-peak periods indicated that their variations were uncorrelated and that there were differences in journey-times of traffic between different sections of road. In order to determine which factors influence the variations of journey-time between sections, an extensive series of roads in the London Boroughs of Barnet, Enfield and Haringey were studied. The "off-peak journey-time" used in this study is the total journey-time less any delay at major intersections. The parameters which were investigated included: traffic flows, parked vehicles, delays at major intersections, variations of land-use, numbers of pedestrian crossings, numbers of bus stops, bus frequency, numbers of priority junctions and road widths. The studies were carried out over 44.5 miles of single-carriageway roads divided into half-mile sections.

## SUMMARY AND CONCLUSIONS

After removing the effects of major intersections it was found that the most significant variables were numbers of pedestrian crossings, the extent of shopping land-use alongside the roads and numbers of bus stops. It was found that variations of journey-time were not correlated with variations of flow between sections. Variations of roadside parking were not found to be significantly correlated with variations of journey-time. Significant relationships were found to exist between a number of variables which had been treated as independent for the purposes of regression analysis and so other methods of multivariate analysis were used. Factor analysis revealed evidence of a significant factor associated with shopping and residential land-uses, and numbers of pedestrian crossings. Variations of this factor were highly correlated with variations of journey-time. This factor can be regarded as an "Activity Index" and is similar to the Activity Index of Buchanan (1976). It has been argued that the "Activity Index" derived using factor analysis has a number of advantages over Buchanan's index. The "Activity Index" derived by factor analysis has been used as the independent variable in a linear regression model using journey-time as the dependent variable, and this model has been shown to give predictions of similar levels of accuracy as those derived by conventional multiple linear regression analysis of the raw variables. Factor models may prove more useful since the theory justifies the use of a wider range of variables than occur in the multiple linear regression models.

Some evidence has been found to indicate that spatial correlations between journey-times and a number of variables can give rise to differences in regression coefficients if investigations are carried

## SUMMARY AND CONCLUSIONS

out on roads of different lengths. Taking into account the earlier observation that changes in time-based sampling intervals can give rise to consistent changes in regression coefficients, these results suggest that some attempt should be made to standardise both time and spatial sampling intervals. Otherwise valid comparisons of the results of different studies may not be possible. The results of the peak-hour studies reported in this thesis show that significant relationships can be obtained using sampling intervals of five minutes and, furthermore, the use of the model reported by Wright (1971) permits the results to be adjusted to larger sampling intervals if these are required. The off-peak studies have shown that significant models of journey-time variations can be obtained using half-mile spatial sampling intervals but, because the models are likely to be used for sections of road of differing lengths, difficulties in the use of such models may occur if spatial cross-correlation exists in the data.

The peak and off-peak studies can be seen to be linked. The off-peak studies have provided a satisfactory method of predicting journey-times, and the corresponding journey-speeds, at periods of low flows and these predictions can be used in other journey-time/flow, or speed/flow, relationships intended for use over a wider range of flows. In any such relationship the off-peak journey-time is usually independent of the traffic flow. It has been shown that these predictions can be used in Davidson's model, thus aiding the model's wider use. This model has the potential to provide a widely applicable predictor of journey-times since it permits the complete range of journey-times to be predicted from the same function.

## SUMMARY AND CONCLUSIONS

Whatever models are used there should be some general agreement on the choice of variables, the units in which they are measured and their sampling intervals.

### 8.3 Suggestions for Further Work.

Additional work is desirable in a number of areas considered in this investigation. Promising results derived from the use of factor analysis on the peak-hour data to produce a peak-hour index suggest that this index should be given further study. In addition, if the equation derived by Davidson (1966) to predict journey-times is to be used more widely then more work needs to be done in order to determine a rational method of measuring the factor J. Wright's model for predicting slope regression coefficients should also be subjected to further investigations, particularly on city centre roads where current speed-flow models are likely to be most inaccurate. On such roads where low average journey-speeds are likely to occur accurate predictions are difficult to obtain because of the relatively wide range of journey-speeds which are found in practice.

Because the data for the off-peak study was collected in three adjacent outer London Boroughs there are similarities in the values of the variables of land-use and the other characteristics. Hence, before wider application of any of the empirical relationships between journey-time and these characteristics, further work needs to be carried out to investigate the nature of such relationships, if any, in some of the inner London Boroughs. The results are likely to be different since journey-time per unit distance tends to be higher in

## SUMMARY AND CONCLUSIONS

the inner London Boroughs. However, this may be largely due to the higher density of major intersections in these areas, in which case the results discussed in this thesis should be applicable to these areas after an allowance has been made for the additional delays at such intersections.

Additional work is necessary to determine whether significant spatial cross-correlation exists between journey-time along a section of road and the characteristics of other nearby sections.

The methods of factor analysis could be applied to other areas of transport study since, as argued in Chapter 5, these methods can deal with the problems of intercorrelation which occur frequently in them. Finally, the methods have the potential to explore the effects of human behaviour in many areas such as trip generation or accident studies.

## APPENDIX A

### REFERENCES

Akcelik R. (1978); A New Look at Davidson's Travel Time Function. Traffic Engineering and Control Vol.19, No.10.

Almond J. (1964); Traffic Assignment to a Road Network with Journey Time/Flow Relations on Each Link. RRL Note LN571. (Unpublished).

Bampfylde A.P., Porter G.J.D., Priest S.D. (1978); Speed/Flow Relationships in Tunnels. Traffic Engineering and Control Vol.19, No.8.

Beard C.L, Maclean J.R. (1974); Speed-Flow Relationship for Nine 4-lane Roads in Melbourne. Proc. Australian Road Research Board, Vol.7, No.4.

Bennett S., Bowers D. (1976); An Introduction to Multivariate Techniques for Social and Behavioural Sciences. Macmillan Press Ltd.

Bilham E.G. (1935); Classification of Heavy Falls in Short Periods. British Rainfall. Vol.75.

Branston D. (1973); Urban Traffic Speeds - 1: A Comparison of Proposed Expressions Relating Journey Speed to Distance from a Town Center. Transportation Science, Vol.7, No 1.

Branston D. (1976); Link Capacity Functions : A Review. Transportation Research. Vol.10, No.4.

Branston D. (1977); A Note on Speed/Flow/Concentration Relations. Traffic Studies Group, University College London. (Unpublished).

Breiman L., Lawrence R.L. (1973); Time Scales, Fluctuations and Constant Flow Periods In Uni-directional Traffic. Transportation Research Vol.7.

Breiman L. (1974); A Relationship Between Space and Time Distributions in Homogeneous Traffic Flow and Its Applications. Transportation Research Vol.8.



## REFERENCES

- Buchanan C.M, Coombe R.D. (1974); An Analysis of Running Speed Data Collected on London Roads. GLC.Dept. of Planning and Transportation, Transportation Branch Background Paper No.20.
- Buchanan C.M, Coombe R.D. (1976); A Note on Traffic Speeds in London. Traffic Engineering and Control, Vol.17, No.6.
- Bureau of Public Roads (1964); Traffic Assignment Manual. U.S. Dept. of Commerce, Urban Planning Division.
- Bureau of Public Roads (1965); Highway Capacity Manual, SR 87. Highway Research Board, Washington.
- Carroll J.B. (1953); An Analytic Solution for Approximating Simple Structure in Factor Analysis. Psychometrika Vol.18.
- Child D. (1970); The Essentials of Factor Analysis. Holt, Rinehart and Winston.
- Chalmers C.P., Galbraith J.I. (1977); Multivariate Data Analysis - Course Manual. Birkbeck College.
- Coleman R.R. (1961); A Study of Urban Travel Times in Pennsylvania Cities. H.R.B. Bulletin 303.
- Comrey A.L. (1973); A First Course in Factor Analysis. Academic Press.
- Cooley W.W, Lohnes P.R. (1971); Multivariate Data Analysis. Wiley.
- Daling J.R, Tamura H. (1970); Use of Orthogonal Factors for Selection of Variables in a Regression Equation. Applied Statistics Vol.19, No.3.
- Davenport A.G. (1964); Note on the Distribution of the Largest Value of a Random Function with Application to Gust Loading. Proc.Inst.Civil Eng. Vol.28, No.6.
- Davidson K.B. (1966); A Flow Travel Time Relationship for Use in Transportation Planning. Proc. Australian Road Research Board, Vol.3, Part 1.
- D.O.E. Highways Directorate (1975); Technical Memorandum on Roundabout Design, H2/75.
- Dick A.C. (1966); Speed/Flow Relationships within an Urban Area. Traffic Engineering and Control, Vol.7, No.10.
- Drake J.S., Schofer J.L., May A.D. (1967); A Statistical Analysis of Speed Density Hypotheses. Highway Research Board, Research Record No. 154.
- Duncan N.C. (1974); Rural Speed/Flow Relations. Transport and Road Research Laboratory Report, LR No.651.

## REFERENCES

- Duncan N.C. (1976); A Note on Speed/Flow/Concentration Relationships. Traffic Engineering and Control, Vol.17, No.1.
- Duncan N.C. (1979); A Further Look at Speed/Flow/Concentration. Traffic Engineering and Control, Vol.20, No.10.
- Edie L.C. (1965); Traffic Stream Measurements and Definitions. Proc. of 2nd Int. Symposium on the Theory of Traffic Flow, O.E.C.D. 1963.
- Evans L., Herman R., Lam T. (1976); Multivariate Analysis of Traffic Factors Related to Fuel Consumption in Urban Driving. Transportation Science, Vol.10, No 2.
- Farouki O.T., Nixon W.J. (1976); The Effect of the Width of Suburban Roads on the Mean Free Speed of Cars. Traffic Engineering and Control, Vol.17, No.12.
- Ford W.G. (1977); Prediction of Vehicle Speeds on Roads in the Lake District. Transport and Road Research Laboratory, Report No SR 230.
- Foster J. (1962); An Investigation of the Hydrodynamic Model for Traffic Flow with Particular Reference to the Effect on Various Speed-Density Relationships. Proc. Australian Road Research Board, Vol.1.
- Fredland D.R. (1976); Multivariate Analysis of Suburban Land-use Clustering. Transportation Planning and Technology, Vol.3 No.2.
- Freeman, Fox and Assoc. (1972); Speed/Flow Relationships on Suburban Main Roads.
- Gipps P.G. (1976); The Interpretation of Vehicle Detection Output by Canonical Variates Analysis. Proc. Seventh Int. Symposium on Transportation and Traffic Theory.
- Goldstein L.G., Mosel J.N. (1958); A Factor Study of Drivers' Attitudes, with Further Study on Driver Aggression. Highway Research Board, Bulletin No.172.
- Gordon R.A. (1968); Issues in Multiple Regression. American Journal of Sociology, Vol.73.
- Greenshields B.D. (1952); Statistics with Application to Highway Traffic Analysis. Eno Foundation for Highway Traffic Control.
- Gudmundsson G. (1977); Multivariate Analysis of Economic Variables. Applied Statistics, Vol.26, No 1.
- Guerin N.S. (1958); Travel Time Relationships. Bureau of Highway Traffic, Yale University.
- Harman H.H. (1967); Modern Factor Analysis. Univ. of Chicago Press.

## REFERENCES

- Harris R.J. (1975); A Primer of Multivariate Statistics. Academic Press
- Harvey A.C. (1977); Some Comments on Multicollinearity in Regression. Applied Statistics, Vol.26, No.2.
- Herman R., Potts R.B. (1961); Single-lane Traffic Theory and Experiment. Theory of Traffic Flow: Elsevier.
- Herman R., Lam T. (1974); Trip Time Characteristics of Journeys to and from Work. Transportation and Traffic Theory: Proc. of Sixth Int. Symposium.
- Horst P. (1965); Factor Analysis of Data Matrices. Holt.
- Irwin N.A., Von Cube H.G. (1962); Capacity Restraint in Multi-Travel Mode Assignment Programmes. Highway Research Board, Bulletin No.347.
- Jolliffe I.T. (1972); Discarding Variables in a Principal Component Analysis 1: Artificial Data. Applied Statistics, Vol.21, No.2.
- Jolliffe I.T. (1973); Discarding Variables in a Principal Component Analysis 2: Real Data. Applied Statistics, Vol.22, No.1.
- Joreskog K.G., Klován J.E., Rayment R.A. (1976); Geological Factor Analysis. Elsevier.
- Kaiser H.F. (1958); The Varimax Criterion for Analytic Rotation in Factor Analysis. Psychometrika, Vol.23.
- Kendall Sir M. (1975); Multivariate Analysis. Griffin.
- Kinloch A.G. (1972); Highway Data Bank Users' Manual. GLC Research Memorandum No.293.
- Large B.J.W. (1971); Highway Data, Derived Capacities and Improvement Costing. GLC Research Memorandum No.227.
- Lawley D.N., Maxwell A.E. (1971); Factor Analysis as a Statistical Method. Butterworths.
- Lynan D.A., Everall P.F. (1971); Car Journey-Times in London - 1970. Transport and Road Research Laboratory, LR No.416.
- Mansfield E.R., Webster J.T., Gunst R.F. (1977); An Analytic Selection Technique for Principal Component Regression. Applied Statistics, Vol.26, No.1.
- Massy W.F. (1965); Principal Component Regression in Exploratory Research. J. American Statistical Assoc., Vol.60.
- Menon A.P.G., Tierney A.J.H., Blunden W.R., Tindall J.I. (1974); A Study of Davidson's Flow Travel Time Relationship. Proc. Australian Road Research Laboratory, Vol.7, No.4.

## REFERENCES

- Meta Systems Inc. (1975); Systema Analysis in Water Resources Planning. Water Information Center Inc.,USA.
- Mosher W.W. (1963); A Capacity Restraint Algorithm for Assigning Flow to a Transport Network. Highway Research Board, Record No.6.
- Neter J., Wasserman W. (1974); Applied Linear Statistical Models. Irwin.
- Newby R.F., Wardrop J.G. (1950); The Effect of Road Width and Flow on the Speed of Traffic in Central London. Road Research Laboratory, RN/1367/RFN,JGW. (Unpublished).
- Newell G.F. (1968); Stochastic Properties of Peak Short-Time Traffic Counts. Transportation Science, Vol.2,No.3.
- Nie N.H., Hull C.H., Jenkins J.G., Steinbrenner K., Bent D.H. (1975); Statistical Package for the Social Sciences: SPSS. McGraw-Hill Book Co.
- O'Muircheartaigh C.A., Payne C. (1977); The Analysis of Survey Data Vol.2: Model Fitting. Wiley and Sons.
- Oppenlander J.C., Bunte W.F., Kadakia P.L. (1961); Sample Size Requirement for Vehicle Speed Studies. Highway Research Board, Bulletin No.287.
- Oppenlander J.C. (1963); Multivariate Analysis of Vehicular Speeds. Highway Research Board, Record No.35.
- Overall J.E. (1962); Orthogonal Factors and Uncorrelated Factor Scores. Psychological Reports, Vol.10.
- Philbrick M.J. (1977a); Conventional Roundabout Capacities - Collection and Analysis of Data from Sites in Sheffield and London. Transport and Road Research Laboratory, Working Paper No.TSN 27R. (Unpublished).
- Philbrick M.J. (1977b); In Search of a New Capacity Formula for Roundabouts. Transport and Road Research Laboratory, LR No.773.
- Road Research Laboratory (1963); Research on Road Traffic. HMSO.
- Rothrock C.A., Keefer L.A. (1957); Measurement of Urban Traffic Congestion. Highway Research Board, Bulletin No.156.
- Ryan D.P., Breuning S.M. (1962); Some Fundamental Relationships of Traffic Flow on a Freeway. Highway Research Board, Bulletin No.324.
- Schonemann P.H. (1966); A Generalised Solution of the Orthogonal Procrustes Problem. Psychometrika, Vol.31.
- Silvey S.D. (1969); Multicollinearity and Imprecise Estimation. Applied Statistics, Vol.18,No.3.

## REFERENCES

- Smeed R.J., Bennett G.T. (1949); Research on Road Safety and Traffic Flow. Proc. Institution of Civil Engineers, Road Division, 1948-1949.
- Smock R. (1962); An Iterative Assignment Approach to Capacity Restraint on Arterial Networks. Highway Research Board, Bulletin No.347.
- Steenbrink P.A. (1974); Optimisation of Transport Networks. Wiley.
- Taylor C.C. (1977); Principal Component and Factor Analysis. Exploring Data Structures; Wiley.
- Taylor M.A.P. (1977); Parameter Estimation and Sensitivity of Parameter Values in a Flow-Rate/Travel-Time Relation. Transportation Science, Vol.11, No.5.
- Thomas E.N., Anderson D.L. (1965); Additional Comments on Weighting Values in Correlation Analysis of Areal Data. Annals of the Assoc. of American Geographers, No.55
- Thurstone L.L. (1965); Multiple Factor Analysis. Univ. of Chicago Press.
- Turner I.K., Wardrop J.G. (1960); The Variation of Journey Time in Central London. Road Research Laboratory, RN/1511. (Unpublished).
- Underwood R.T. (1960); Speed, Volume and Density Relationships. Bureau of Highway Traffic, Yale University.
- Versace J. (1960); Factor Analysis of Roadway and Accident Data. Highway Research Board, Bulletin No.240.
- Wagner F.A., May A.D. (1961); Volume and Speed Characteristics at Seven Study Locations. Highway Research Board, Bulletin No.281.
- Walker W.P. (1957); Speed and Travel Time Measurement in Urban Areas. Highway Research Board, Bulletin No.156.
- Wallis J.R. (1965); Multivariate Statistical Methods in Hydrology. Water Resources Research, Vol.1, No.4.
- Wardrop J.G. (1948); The Relationship Between Speed and Flow in Central London. Road Research Laboratory, Rn/1021/JGW. (Unpublished).
- Wardrop J.G. (1952); Some Theoretical Aspects of Road Traffic Research. Proc. Institution of Civil Engineers. Vol.1, No.2.
- Wardrop J.G., Duff J.T. (1956); Factors Affecting Road Capacity. Proc. Third Int. Study Week in Traffic Engineering (Stresa).
- Wardrop J.G. (1957); The Traffic Capacity of Weaving Sections of Roundabouts. Proc. First Int. Conference on Operational Research, (Oxford), English Universities Press Ltd.

## REFERENCES

Wardrop J.G. (1963); Experimental Speed/Flow Relations in a Single Lane. Proc. Second Int. Symposium on the Theory of Traffic Flow. O.E.C.D.

Wardrop J.G. (1968); Journey Speed and Flow in Central Urban Areas. Traffic Engineering and Control, Vol.9.

Weiner S. (1974); A Review of the Traffic Flow Process. Transportation Research Board, Transportation Research Record No 509.

Wortman R.H. (1965); A Multivariate Analysis of Vehicular Speeds on Four-Lane Rural Highways. Highway Research Board, Record No.72.

Wright C.C. (1971); Characteristics of Single-Lane Traffic Flow. Ph.D Thesis, CNAA.

Wright C.C., Hyde T. (1972); The Measurement and Interpretation of Speed-Flow Relationships. Traffic Engineering and Control, Vol.13, No.3.

APPENDIX B

TABLES

Table 3.1

Journey Time/Flow Study, Survey Details.

Location	Date	No. of 1 Minute Intervals	Mean Flow (veh/min)	Mean J. Time (min)	J. Time Sample (%)
High Street S/Bound	17. 7.72	90	8.87	1.09	17
Ditto	19. 7.72	90	9.78	1.76	24
Ditto	13.11.72	32	8.10	1.11	71
Ditto	22.10.73	42	9.40	2.38	83
Southbury Rd. E/Bd.	19.10.73	80	15.80	1.43	45
Ditto W/Bd.	Ditto	78	24.82	1.58	71
Ditto E/Bd.	5.12.73	73	17.95	1.92	60
Ditto W/Bd.	Ditto	77	14.49	2.19	55

(Continued overleaf)



Table 3.1 Cont/...

Journey Time/Flow Study, Survey Details

Location	Date	Methods	Other Variables	Remarks
High Street S/Bound	17. 7.72	Reg. Nos. Manual & Automatic counts	Vehicle types Turning vehs. Crossing peds. Opposing flow Green times	Off-peak
Ditto	19. 7.72	Ditto	Ditto	Peak
Ditto	13.11.72	Cine-photo- graphy Manual counts	Vehicle types Opposing flow	Off-peak
Ditto	22.10.73	Ditto	Ditto	Peak
Southbury Rd. E/Bd.	19.10.73	Ditto	Ditto	p.m. Peak
Ditto W/Bd.	Ditto	Ditto	Ditto	Ditto
Ditto E/Bd.	5.12.73	Ditto	Ditto	a.m. Peak
Ditto W/Bd.	Ditto	Ditto	Ditto	Ditto (Includes eff- ects of petrol station queues)

Table 3.2

Linear Regression Analyses of High Street Data showing the Correlation and Slope Regression Coefficients for Sampling Periods of 1 Minute and 5 Minutes

Date	Sample Size 1 Minute Intervals	Mean J.Time (min)	Mean Flow (Veh/min)	Corrn. Coeff		Slope Regn Coeff		Remarks
				1 min	5 min	1 min	5 min	
17. 7.72.	90	1.09	8.87	.110	.351	.012	.204	Off-peak
19. 7.72.	90	1.76	9.78	.251*	.666*	.051	.230	Peak
13.11.72.	32	1.11	8.10	.009	-.405	.001	-.020	Off-peak
22.10.73.	42	2.38	9.40	.306*	.686*	.095	.272	Peak

\* Significantly greater than zero.

The units of the Slope Regression Coefficient are min/veh/min

Table 3.3

Slope Regression and Correlation Coefficients of the Relationships between Flow and Flow-Weighted/Unweighted Journey-Times for Southbury Road, (westbound) 19.10.73.

Period of Aggregation (minutes)	Weighted Models		Unweighted Models	
	Slope	Corr. Coeff.	Slope	Corr. Coeff.
1	.033	.352	.033	.352
2	.064	.485	.065	.496
3	.073	.521	.076	.535
4	.097	.571	.100	.587
5	.099	.652	.104	.672
6	.125	.674	.126	.681
7	.128	.664	.132	.682
8	.153	.680	.152	.683
9	.135	.766	.140	.796
10	.156	.694	.167	.736

The units of slope are min/veh/min.

Table 3.4

Effects of Changing Sampling Intervals on Regression Equations  
between Journey-time and Traffic Flow.

High Street 19.7.72.			
Interval of Average (minutes)	No. of Data	Correlation Coeff.	Regression Equation
1	83	.251	$Y = 1.253 + .051X$
2	44	.403	$Y = .638 + .116X$
3	30	.506	$Y = .126 + .168X$
4	22	.554	$Y = -.025 + .182X$
5	18	.665	$Y = -.498 + .230X$
6	15	.651	$Y = -.582 + .240X$
8	11	.604	$Y = -.470 + .227X$
10	9	.767	$Y = -1.399 + .322X$
High Street 22.10.73.			
Interval of Average (minutes)	No. of Data	Correlation Coeff.	Regression Equation
1	42	.306	$Y = 1.486 + .095X$
2	21	.428	$Y = .923 + .152X$
3	14	.562	$Y = .126 + .270X$
4	10	.541	$Y = .324 + .212X$
5	8	.686	$Y = .228 + .272X$
6	7	.745	$Y = -1.717 + .446X$
7	6	.761	$Y = .880 + .353X$
8	5	.914	$Y = -1.876 + .453X$
10	4	.826	$Y = -.285 + .278X$

Table 3.5

## Linear Regression Analyses of Southbury Road Data

Date	Sample Size 1 minute Intervals	Mean J.Time (min)	Mean Flow (v/min)	J. Time Sample %	Correlation Coeff.		Remarks
					1 min	5 min	
19.10.73.	80	1.26	25.18	45.6	.097	- .002	E.Bound p.m.peak
19.10.73.	58	1.57	25.20	72	.354*	.600*	W.Bound p.m.peak
5.12.73.	73	1.92	17.95	60	.412*	.737*	E.Bound a.m.peak
5.12.73.	77	2.19	18.00	55	- .136	- .353	W.Bound a.m.peak

\* Significantly greater than zero.

Notes: See Table 3.2

Table 3.6

Correlation Matrix of Variables Used in Multiple Linear Regression Analysis - High Street, 19.7.72. (5 minute intervals)

Variables	X1	X6	X7	X8	Y
Traffic Flow (X1)	1.000	.539	.668	.128	.666
Pedestrians Crossing (X6)		1.000	.780	-.474	.521
Turning Vehicles (X7)			1.000	-.126	.653
Opposing Flow (X8)				1.000	-.111
Journey-time (Y)					1.000

Notes: Sample size 18

Correlation coefficients numerically greater than .41 are significant at the 5% level.

Table 3.7

Multiple Linear Regression Results - High Street, 19.7.72.

Independent Var.	Multiple R	Regression Equation
X1, X6, X7, X8	.739	$Y = .500 + .174 X1 - .040 X6 + .169 X7 - .072 X8$
X1, X6, X7	.722	$Y = -.428 + .143 X1 - .002 X6 + .140 X7$
X1, X6, X8	.699	$Y = .255 + .211 X1 + .026 X6 - .042 X8$
X1, X7, X8	.732	$Y = .212 + .160 X1 + .124 X7 - .043 X8$
X6, X7, X8	.654	$Y = .554 + .0002 X6 + .247 X7 - .010 X8$
X1, X6	.693	$Y = -.313 + .187 X1 + .045 X6$
X1, X7	.722	$Y = -.435 + .143 X1 + .143 X7$
X1, X8	.694	$Y = .513 + .239 X1 - .067 X8$
X6, X7	.653	$Y = .405 + .006 X6 + .239 X7$
X6, X8	.543	$Y = .178 + .120 X6 + .958 X8$
X7, X8	.654	$Y = .556 + .246 X7 - .010 X8$
X1	.665	$Y = -.498 + .230 X1$ *
X6	.521	$Y = 1.225 + .103 X6$ *
X7	.653	$Y = .387 + .247 X7$ *
X8	-.111	$Y = 2.361 - .372 X8$

\* These models are significant at the 5% level.

Table 3.8

Lagged Serial and Cross-Correlations between Flow and Journey-Time, High Street, 19.7.72. (Peak Period)

	Lag (intervals)									
	0	1	2	3	4	5	6	7	8	9
Flow and Lagged J.T.	1.000	-.074	.116	.165	-.012	<u>.285</u>	.036	.215	-.004	-.094
Flow and Lagged Flow	<u>.251</u>	<u>.262</u>	.140	.171	.174	.017	.046	.072	.052	-.141
Journey Time and Lagged J.T.	1.000	<u>.800</u>	<u>.670</u>	<u>.552</u>	<u>.490</u>	<u>.377</u>	<u>.252</u>	.184	.154	.067
Journey Time and Lagged Flow	<u>.251</u>	<u>.316</u>	<u>.388</u>	<u>.343</u>	<u>.363</u>	<u>.365</u>	<u>.437</u>	<u>.308</u>	<u>.298</u>	.153

The underlined coefficients are significantly different from zero at the 5% level.



Table 3.9

Slope Regression and Correlation Coefficients Predicted by Wright's Model Compared with Values Calculated from the Aggregated Data Southbury Road, (westbound) 19.10.73.

Period of Aggregation (minutes)	Coefficients of Aggregated Data		Predicted Coefficients	
	Slope	Correlation	Slope	Correlation
1	.033	.352	.032	.352
2	.064	.485	.062	.483
3	.073	.521	.082	.553
4	.097	.571	.094	.584
5	.099	.652	.104	.632
6	.125	.674	.120	.667
7	.128	.664	.133	.709
8	.153	.680	.143	.740
9	.135	.766	.150	.761
10	.156	.694	.159	.787

The units of slope are min/veh/min.

Table 3.10

Maximum Values of Slope Regression and Correlation Coefficients  
Predicted by Wright's Models. (All Peak Period Data)

Sample		Max. Predicted Coefficient		Optimum Lag (Slope)
		Slope	Correlation	
High Street	19. 7.72	.384	.977	36
High Street	22.10.73	.583	*1.015	40
Southbury Rd (E)	19.10.73	NA	NA	NA
Southbury Rd (W)	19.10.73	.169	.837	14
Southbury Rd (E)	5.12.73	.420	.999	41
Southbury Rd (W)	5.12.73	NA	NA	NA

Notes:

1. The results for Southbury Road and marked NA were not amenable to analysis because of the lack of linear correlation between the variables.
2. \* This unacceptable result is discussed in the text.
3. The data collected on the High Street on 17.7.72. and 13.11.72. are not included because the data was collected during off-peak periods.
4. The units of slope are min/veh/min.

Table 3.11

Maximum Slope and Correlation Coefficients Predicted by Wright's Models. High Street, 19.7.72.

Basic Sampling Interval minutes	Optimum Lag		Maximum Slope	Correlation Coefficient
	Intervals	Minutes		
1	36	36	.384	.975
2	17	34	.377	.957
3	12	36	.381	.966
4	9	36	.407	.974
5	7	35	.377	.966
6	6	36	.390	.975
8	5	40	.399	.962
10	4	40	.365	.953

The units of slope are min/veh/min.

Table 3.12

Relationships between Journey-Times and Pseudo-Concentration,  
5 minute averages.

Location	Apparent Coefficients		Actual Coefficients	
	Correlation	F	Correlation	F
High St. 17. 7.72	.909	71.5	.191	.57
ditto 19. 7.72	.956	169.7	.408	3.19
ditto 13.11.72	.635	2.7	-.238	.24
ditto 19.10.73	.973	105.3	.383	1.03
Southbury Rd. 19.10.73 Eastbound	.932	92.3	.041	.02
Westbound	.976	256.3	.725	14.38
Southbury Rd. 5.12.73 Eastbound	.965	162.8	.755	15.86
Westbound	.897	53.3	.245	.83

The apparent coefficients were determined for the model

$$T = A_0 + A_1 (QT)$$

The actual coefficients were determined for the predictive model

$$T = \frac{A_0}{1 - A_1 Q}$$

Table 4.1

Combined Sum of Squares of Discontinuous Relationship between Journey-Time and Flow (5 Minute Averages), High Street, Total Data.

Critical Flow (veh/min)	Combined Sum of Squares (min <sup>2</sup> )	Mean Square Error of predicted journey-time
7.0	21.11	.449
7.5	19.35	.412
8.0	18.20	.387
8.5	17.61	.375
9.0	18.09	.385
9.5	20.24	.431
10.0	21.40	.455
10.5	27.55	.586

Table 4.2

Combined Sum of Squares and Residual Squares of Relationships between Journey-Time and Mean Flow, High Street, Total Data.

K (min- utes)	Discontinuous Model		Least-Squares Model	
	Minimum Mean Square Error	Optimum Slope	Mean Square Error	Regression Slope
1	.663	.056	.664	.056
2	.575	.143	.577	.120
3	.503	.225	.520	.175
4	.427	.228	.447	.193
5	.375	.268	.394	.237
6	.380	.438	.410	.282
10	.169	.683	.289	.323

Note: The slopes are expressed in units of minutes/vehicle/minute.

Table 4.3

Combined Sum of Squares and Residual Squares of Relationships between Mean Flow and Journey-Speed, High Street, Total Data.

K (minutes)	Discontinuous Model		Least-Squares Model	
	Minimum Mean Square Error	Optimum Slope	Mean Square Error	Regression Slope
1	28.8	No Optm	28.8	- .252
2	23.3	- .828	23.3	- .630
3	20.4	- 1.35	21.4	- .990
4	18.5	- 2.18	19.8	-1.134
5	16.8	- 1.28	17.2	-1.40
6	17.2	- 2.21	18.1	-1.66
10	9.07	- 4.05	13.4	-1.76

Note: The slopes are expressed in mph/veh/minute.

Table 4.4

Discontinuous Regression using a Power Curve on High Street Combined Data. Relationship between Journey-Time and Flow.

K (minutes)	A	Vo	Optimum Qk	Mean Square Error
1	-.370	.812	10.5	.600
2	.000	.813	10.5	.545
3	.104	.968	10.5	.468
4	.380	.577	9.00	.388
5	-.129	1.039	10.50	.301
6	-.113	1.561	10.00	.179
8	-.274	1.077	11.00	.258
10	-.017	1.359	10.50	.118

Notes: Qk = Critical Flow (veh/min)  
A = Power of the curve  
Vo = Intercept

Table 4.5

Analysis of High Street Total Data using Davidson's Travel Time Model to Predict Journey-Times.

K (minutes)	Optimum Values			Mean Square Error
	S (veh/min)	J	t (min)	
1	28.5	.356	1.30	.666
2	22.0	.681	1.00	.575
3	20.5	.818	.90	.515
4	20.0	.772	.90	.440
5	19.5	.837	.85	.425
8	17.5	.980	.70	.328
10	15.0	1.313	.45	.213

Notes: S = saturation flow, t = travel time at zero flow  
J = Davidson's constant

Table 4.6

Saturation Flows - High Street using Davidson's, Poisson and Normal Distributions.

K (minutes)	Flow parameters		Saturation Flows (veh/min)			
	Mean (veh/int)	Variance	Davidson	Poisson	Normal (1)	Normal (2)
1	9.39	20.03	28.5	19.0	22.1	25.2
2	18.50	40.04	22.0	15.3	17.4	20.2
3	27.80	60.05	20.5	13.7	15.5	18.5
4	36.85	80.07	20.0	12.9	14.4	17.6
5	46.22	100.12	19.5	12.4	13.7	17.0
6	55.66	120.14	19.0	12.1	13.2	16.6
10	92.54	200.22	15.0	11.2	12.0	15.5

Table 5.1

Correlation Matrix of Data Collected on 17.7.72. in the High Street.  
(Off peak)

X1 = Southbound traffic flow  
X2 = Numbers of pedestrians crossing  
X3 = Numbers of turning vehicles  
X4 = Northbound traffic flow  
Y = Mean journey-time (southbound)

	X1	X2	X3	X4	Y
X1	1	-.059	.343	.194	.223
X2		1	.028	.236	.034
X3			1	-.106	.206
X4				1	-.268
Y					1

Table 5.2

Correlation Matrix of Data Collected on 19.7.72 in the High Street  
(Peak Period)

	X1	X2	X3	X4	Y
X1	1	.539	.668	.128	.665
X2		1	.780	-.474	.521
X3			1	-.126	.653
X4				1	-.111
Y					1



Table 5.3

Eigenvalues and Percentage of Variation Explained by the Principal Components of Variables X1, X2, X3 and X4: High Street.

Principal Component	17.7.72.		
	Eigenvalue	% age of var.	Cumulative
1	1.359	34.0	34.0
2	1.247	31.2	65.2
3	.895	22.4	87.6
4	.498	12.5	100.0

Principal Component	19.7.72.		
	Eigenvalue	% age of var.	Cumulative
1	2.390	59.8	59.8
2	1.176	29.4	89.2
3	.296	7.4	96.6
4	.137	3.4	100.0

Table 5.4

Loadings on Two Principal Components in the Peak Period, High Street,  
19.7.72. 5 Minute Data, N = 18.

Variable	Principal Components		Rotated Factors		Communality
	F1	F2	F1'	F2'	
X1	.768	.501	.891	.216	.841
X2	.924	-.246	.788	-.541	.914
X3	.917	.155	.916	-.160	.864
X4	-.327	.917	-.002	.973	.947

Table 5.5

Effects of Data Aggregation into Successively Longer Sampling Intervals  
on the Eigenvalues of the Principal Components, High Street, 19.7.72.

K	Eigenvalues				Number of Data
	F1	F2	F3	F4	
1	1.52	1.12	.82	.43	83
2	1.87	1.20	.59	.34	44
3	2.04	1.20	.52	.25	30
4	2.11	1.15	.49	.24	22
5	2.39	1.18	.30	.14	18
6	2.37	1.04	.37	.22	15
7	2.20	1.19	.54	.07	12
8	2.31	1.25	.38	.06	11

Table 6.1

Parameters Measured in Journey-Time Study, Enfield (1979)  
(Single-Carriageways Only) - 45 Sections.

Parameter	Var. Label	Units	Mean	Minimum	Maximum
Road Width	X2	Feet	34.46	20.0	48.12
Bus Frequency	X3	Buses/Hr	11.40	3	24
LAND-USE					
Residential	X4	%	56.2	0	100
Shopping	X5	%	26.0	0	100
Other		%	17.8	0	90
Pedestrian Crossings	XS	No.	1.27	0	4
Side-Turnings	XT	No.	6.14	1	12
Bus-Stops	XB	No.	4.29	0	7
Off-Peak Traffic Flow	X11	Veh/Hour	1076	445	1760
Parked Vehicles	X12	No.	18.3	0	147
Running Journey-Time		Minutes	1.22	0.72	1.98

Notes: The variable labels apply to the corresponding variables in  
• Tables 6.2 and 6.3 and are used in the models given in  
Tables 6.4 to 6.15.

Table 6.2

Parameters Measured in Journey-Time Study, Haringey (1979)  
(Single-Carriageways Only) - 15 Sections.

Parameter	Units	Mean	Minimum	Maximum
Road Width	Feet	34.60	27.56	41.34
Bus Frequency	Buses/Hr	14.20	2.00	34.00
LAND-USE				
Residential	%	45.87	0	90
Shopping	%	25.0	0	83
Other	%	29.13	2	70
Pedestrian Crossings	No.	1.00	0	3
Side Turnings	No.	6.87	4	11
Bus-Stops	No.	4.60	1	11
Off-Peak Traffic Flow	Veh/Hour	1148	697	2000
Parked Vehicles	No.	30.3	1	141
Running Journey-Time	Minutes	1.37	0.92	1.84

Table 6.3

Parameters Measured in Journey-Time Study, Barnet (1980)  
(Single-Carriageways Only) - 39 Sections.

Parameter	Units	Mean	Minimum	Maximum
Road Width	Feet	33.55	20.0	50.0
Bus Frequency	Buses/Hr	16.4	0	34.0
LAND-USE				
Residential	%	52.4	10	100
Shopping	%	22.9	0	80
Other	%	25.3	0	75
Pedestrian Crossings	No.	1.00	0	3
Side Turnings	No.	8.97	2	16
Bus-Stops	No.	4.46	0	7
Off-Peak Traffic Flow	Veh/Hour	979	388	1906
Parked Vehicles	No.	48.3	0	152
Running Journey-Time	Minutes	1.20	0.88	1.66

Table 6.4

Results of Linear Regression Analysis of Data Collected on Dual-Carriageways in Enfield - 1978 (Sample Size - 15)

Dependent Variable	Model	Multiple R	Comments
Journey-time (Y) (minutes)	$Y = .615 + .021X_3$	.485	Not Significant
Running-speed (Z) (mph)	$Z = 46.0 - .866X_3$	.447	Not Significant

Table 6.5

Regression Models: Single Carriageway Data (N = 45)

Dependent Variable	Model	Multiple R	Comments
Journey-Time (Y) (min)	$Y = 1.09 + .178X_S$	.736	Highly Significant (a)
	$Y = 1.27 + .145X_S - .0027X_4$	.792	Significant at 1% level (b)
	$Y = 1.10 + .138X_S - .0028X_4 + .043X_B$	.822	Significant (c) at 2.5% level
	$Y = 1.12 + .158X_S - .0029X_4 + .063X_B - .010X_3$ (a)	.843	Significant at 5% level (d)
Running-Speed (Z) (mph)	$Z = 27.8 - 3.08X_S$	.686	Highly Significant
	$Z = 24.6 - 2.51X_S + .046X_4$	.738	Significant at 2.5% level
	$Z = 28.0 - 2.35X_S + .049X_4 - .869X_B$ (b)	.777	Significant at 2.5% level

Table 6.6

Regression Models: Single Carriageway Data (N = 42), Enfield 1979

Dependent Variable	Model	Multiple R	Comments
Journey-Time (Y) (min)	$Y = 1.54 - .0054X_4$	.578	Highly Significant
	$Y = 1.05 + .153X_S$	.567	Highly Significant
	$Y = 1.33 + .103X_S - .0038X_4$	.780	Significant at 5% level
Running-Speed (Z) (mph)	$Z = 29.3 - 2.95X_S$	.527	Highly Significant
	$Z = 24.8 - 2.13X_S + .062X_4$	.597	Significant at 5% level
	$Z = 30.0 - 1.90X_S + .068X_4 - 1.35X_B$	.667	Significant at 5% level

Table 6.7

Correlation Between Observed and Predicted Journey-Times and Running-Speeds, Using Models Derived from the 1978 Data on the 1979 Data Collected in Enfield.

Model	Correlation Coefficient
Predicted Journey-Time = $1.10 + .138X_S - .0028X_4 + .043X_B$	.687
Predicted Running-Speed = $28.0 - 2.35X_S + .049X_4 - .869X_B$	.654

Table 6.8

Regression Models: Single Carriageway Data, Haringey 1979, (N = 15)

Dependent Variable	Model	Multiple R	Comments
Running-Time (Y) (min)	$Y = 1.0 + .022X_3$	.806	Highly Significant
	$Y = 1.07 + .014X_3 + .004X_5$	.844	Significant at 5% level
Running-Speed	$Y = 26.17 - .135X_5$	.718	Highly Significant
	$Y = 12.5 - .160X_5 + .413X_2$	.780	

Table 6.9

Comparison of Models Produced by Enfield 1978, 1979 and Haringey 1979 Data.

Dependent Variable	Model	Multiple R	Comments
Running-Time (Y) (min)	$Y = 1.09 + .178X_S$	.736	Enfield 1978 (N = 45)
	$Y = 1.05 + .153X_S$	.567	Enfield 1979 (N = 42)
	$Y = 1.20 + .155X_S$	.449	Haringey 1979 (N = 15) *
Running-Speed (Y) (mph)	$Y = 27.8 - 3.08X_S$	.686	Enfield 1978 (N = 45)
	$Y = 29.3 - 2.95X_S$	.527	Enfield 1979 (N = 42)
	$Y = 26.3 - 3.14X_S$	.516	Haringey 1979 (N = 15)

\* Significant at 10% level.



Table 6.10

Correlation Between Observed and Predicted Journey-Times and Running-Speeds, Haringey 1979, Using Models Derived From Enfield (1978) Data.

Model		Correlation Coefficient
Predicted Journey-Time	= $1.10 + .138XS - .0028X4 + .043XB$	.601
Predicted Running-Speed	= $28.0 - 2.35XS + .049X4 - .869XB$	.565

Table 6.11

Regression Models: Single Carriageway Data Barnet 1980 (N = 39)

Dependent Variable	Model	Multiple R	Comments
Journey-Time (Y) (min)	$Y = 1.05 + .0063X5$	.696	Significant at 1% level
	$Y = 1.02 + .0052X5 + .057XS$	.726	Significant at 5% level
	$Y = .862 + .0065X5 + .069XS + .002X4$	.748	Ditto
	$Y = .936 + .0050X5 + .053XS + .021XB$	.744	
	$Y = 1.01 + .0038X5 + .0021X12$	.750	Significant at 1% level
	$Y = 1.354 + .0034X5 + .0038X12 - .012X2$	.792	Ditto
Journey-Speed (Y) (mph)	$Y = 29.6 - .077X12$	.674	Significant at 1% level
	$Y = 22.4 - .111X12 + .265X2$	.723	Ditto
	$Y = 23.2 - .084X12 + .240X2 - .056X5$	.751	Significant at 5% level

Table 6.12

Correlation Matrix - Barnet Off-Peak Journey-Time Study 1980.

	Journey-Time	Journey-Speed	X2	X3	X4	X5	XT	XS	XB	X11	X12
Journey-Time	1.000	-.984	.326	.351	-.379	.696	.505	.541	.279	.280	.691
Journey-Speed		1.000	-.330	-.355	.321	-.644	-.484	-.492	-.269	-.250	-.674
X2			1.000	.694	-.388	.478	.448	.297	.415	.724	.748
X3				1.000	-.236	.486	.433	.285	.381	.669	.596
X4					1.000	-.675	-.228	-.509	-.053	-.327	-.523
X5						1.000	.568	.526	.144	.464	.714
XT							1.000	.347	.477	.289	.569
XS								1.000	.145	.339	.516
XB									1.000	.298	.317
X11										1.000	.500
X12											1.000

Table 6.13

Linear Regression Models, Barnet (1980) Using Variables XS, X4 and XB only.

Dependent Variable	Model	Correlation Coefficient	Remarks
Journey-time (Y) (min)	$Y = 1.07 + .127XS$	.541	Significant at 5% level
	$Y = 1.16 + .110XS - .0013X4$	.554	
Journey-speed (Y) (mph)	$Y = 28.4 - 2.46XS$	.492	Significant at 5% level
	$Y = 27.2 - 2.22XS + .019X4$	.499	

Table 6.14

Correlation Between Observed and Predicted Journey-Times and Running-Speeds, Barnet 1980 Using Models Derived from Enfield (1978) Data.

Model		Correlation Coefficient
Predicted Journey-Time	$= 1.10 + .138XS - .0028X4 + .043XB$	.589
Predicted Running-Speed	$= 28.0 - 2.35XS + .049X4 - .869XB$	.533

Table 6.15

Regression Models of Spatially Aggregated Journey-Time Data. Showing Effects of Aggregation on the Coefficients. (Pooled Data - N = 99)

No of Aggregated Sections	Model	Correlation Coefficient
1	$Y = 1.09 + .159XS$	.619
	$Y = 1.06 + .106XS + .0037X^5$	.716
2	$Y = 1.03 + .214XS$	.702
	$Y = 1.01 + .149XS + .0038X^5$	.759
3	$Y = 1.10 + .157XS$	.632
	$Y = 1.07 + .098XS + .0038X^5$	.694
4	$Y = 1.05 + .198XS$	.678
	$Y = 1.01 + .143XS + .0039X^5$	.724
5	$Y = 1.05 + .207XS$	.789
	$Y = 1.03 + .185XS + .0016X^5$	.795

Table 7.1.

Factor Loadings Obtained by Principal Component Analysis and Factor Analysis of the Pooled Off-Peak Journey-Time Data.

Variable	Three Rotated Principal Components			Rotated Factors Derived by Factor Method		
	F1	F2	F3	F1 '	F2 '	F3 '
X2	.068	.759	.181	.099	.710	.059
X3	.295	.680	.345	.310	.694	.198
X4	-.847	-.101	.044	-.760	-.119	.081
X5	.839	.183	.186	.841	.212	.095
XT	.008	.237	.838	.060	.287	.954
XS	.736	.036	.248	.617	.108	.177
XB	-.053	.674	.300	.060	.495	.231
X11	.479	.633	-.178	.411	.393	.026
X12	.362	.225	.676	.345	.401	.290

The following variable labels apply to Tables 7.1 to 7.8:

- X2 = Road Width
- X3 = Off-Peak Bus Frequency
- X4 = Residential Land-use
- X5 = Shopping Land-use
- X7 = Side-Turnings
- XS = Pedestrian Crossings
- XB = Bus Stops
- X11 = Off-Peak Traffic Flow
- X12 = Parked Vehicles

Table 7.2.

Correlation Matrix of Rotated Factors and Journey-Times for the Pooled Off-Peak Data. Factors Derived from Principal Components.

BJT = Normalised Journey-Time

BJS = Normalised Journey-Speed

Variable	BJT	BJS	F1	F2	F3
BJT		-.976	.639	.196	.118
BJS	-.976		-.582	-.185	-.166
F1	.639	-.582		0	0
F2	.196	-.185	0		0
F3	.118	-.166	0	0	

Table 7.3.

Most Significant Linear Regression Equations Between Factor Scores and Off-Peak Journey-Times/Journey-Speeds. Pooled Data.

Models (a) and (c) used Factor Scores derived from Principal Components.

Models (b) and (d) used Factor Scores derived from Common Factor Model.

Dependent Variable	Model	Correlation Coefficient	Remarks
Journey-Time (Y)	$Y = 1.26 + .169 \text{ FS1(a)}$	.639	Significant at 1% level
Journey-Time (Y)	$Y = 1.26 + .191 \text{ FS1 (b)}$	.655	Significant at 1% level
Journey-Speed (Z)	$Z = 24.8 - 2.96 \text{ FS1(c)}$	-.582	Significant at 1% level
Journey-Speed (Z)	$Z = 24.8 - .305 \text{ FS1 (d)}$	-.600	Significant at 1% level

Table 7.4.

Variables Retained by Variable Elimination Processes Types B1 and B2, and Results of Multiple Linear Regression Analysis of the Remaining Variables.

Method	Principal Component Analysis			Multiple Regression Analysis	
Method	Var. Retained	Factor	Eigen-values	Model	Corr. Coeff.
B1	X4	F1	1.00	$Y = 1.41 + .0119XT$	.525
	XT	F2	.99	$- .0045X4$	
B2	X4	F1	1.29	$Y = 1.37 + .0669X3$	.529
	X3	F2	.71	$- .0039X4$	

Y = Journey-time (minutes)



Table 7.5.

Multiple Correlation Coefficients of Each Variable with Remaining Variables, Variable Elimination Process Type A2 - Off-Peak Journey-Time Study, Total Data.

Variable Label \ No. of Variables	9	8	7	6	5	4	3	2
X2	.601	.601	.538	.449	.443	.378	.297	*
X3	.692	.683	*	-	-	-	-	-
X4	.702	.587	.579	.516	.369	.161	.147	.005
X5	.758	*	-	-	-	-	-	-
XS	.606	.556	.547	.543	*	-	-	-
XB	.525	.522	.464	.464	.464	.447	*	-
XT	.584	.583	.568	.466	.409	.405	.264	.005
X11	.558	.548	.537	.530	.494	*	-	-
X12	.633	.621	.615	*	-	-	-	-

\* Indicates variable eliminated at previous stage using maximum multiple correlation coefficient criterion. Coefficients greater than .20 are significant at 5% level. Multiple regression equation of journey-time with X4 and XT.  $Y = 1.43 - .0045X_4 + .012XT$ .  $R = .524$

Table 7.6

Communalities of Nine Variables Based on the Factor Loadings for Two Rotated Factors: Pooled Data.

Variable	Number of Variables						
	9	8	7	6	5	4	3
X4	.727	.740	.750	.750	.835	.817	.834
X5	.765	.785	.785	.790	.813	.825	.828
XS	.557	.565	.552	.583	*	-	-
X11	.410	*	-	-	-	-	-
X12	.471	.517	.505	*	-	-	-
XB	.510	.510	.566	.637	.630	.819	.998
X2	.505	.498	*	-	-	-	-
X3	.649	.649	.616	.645	.673	.723	*
XT	.506	.510	.640	.613	.623	*	-

\* Indicates variable eliminated at the previous stage. The variable having the lowest communality is eliminated.

Table 7.7

Variation of Factor Loadings of First Two Factors as the Number of Variables is Reduced: Pooled Data.

Variable	Number of Variables Retained											
	9		8		7		6		5		4	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
X4	-.853	.015	-.860	-.005	-.866	.025	-.865	.043	-.913	.026	-.903	-.044
X5	.853	.194	.860	.214	.864	.196	.870	.180	.879	.200	.885	.200
XS	.735	.127	.738	.144	.720	.182	.737	.197	-	-	-	-
X11	.542	.340	-	-	-	-	-	-	-	-	-	-
X12	.380	.571	.414	.587	.434	.563	-	-	-	-	-	-
XB	.016	.714	-.011	.714	-.005	.752	.028	.797	.045	.793	-.040	.904
X2	.145	.696	.114	.696	-	-	-	-	-	-	-	-
X3	.362	.720	.344	.729	.371	.692	.393	.700	.417	.706	.367	.767
XT	.029	.711	.038	.713	.026	.800	.040	.782	-.020	.789	-	-

Notes: Factor loadings greater than .28 or less than -.28 are significantly different from zero at the 99% level.

Table 7.8

Details of Multiple Regression Models of Normalised Journey-Time and Factor Scores of Reduced Variable Models: Pooled Data.

Number of Variables	Model	R	F
9	$BJT = .655F1 + .177F2$	.678	40.9
8	$BJT = .642F1 + .188F2$	.669	38.8
7	$BJT = .635F1 + .213F2$	.670	39.0
6	$BJT = .650F1 + .225F2$	.688	43.0
5	$BJT = .580F1 + .236F2$	.627	31.0
4	$BJT = .561F1 + .282F2$	.628	31.3
3	$BJT = .592F1 + .251F2$	.643	33.8

Notes: BJT = Normalised journey-time.

F1 = Factor scores of factor associated with variations of land-use and pedestrian crossings (X4, X5 and XS).

F2 = Factor scores of factor associated with variations of bus frequency, bus stops and parked vehicles (X3, XB and X12).

APPENDIX C

FIGURES

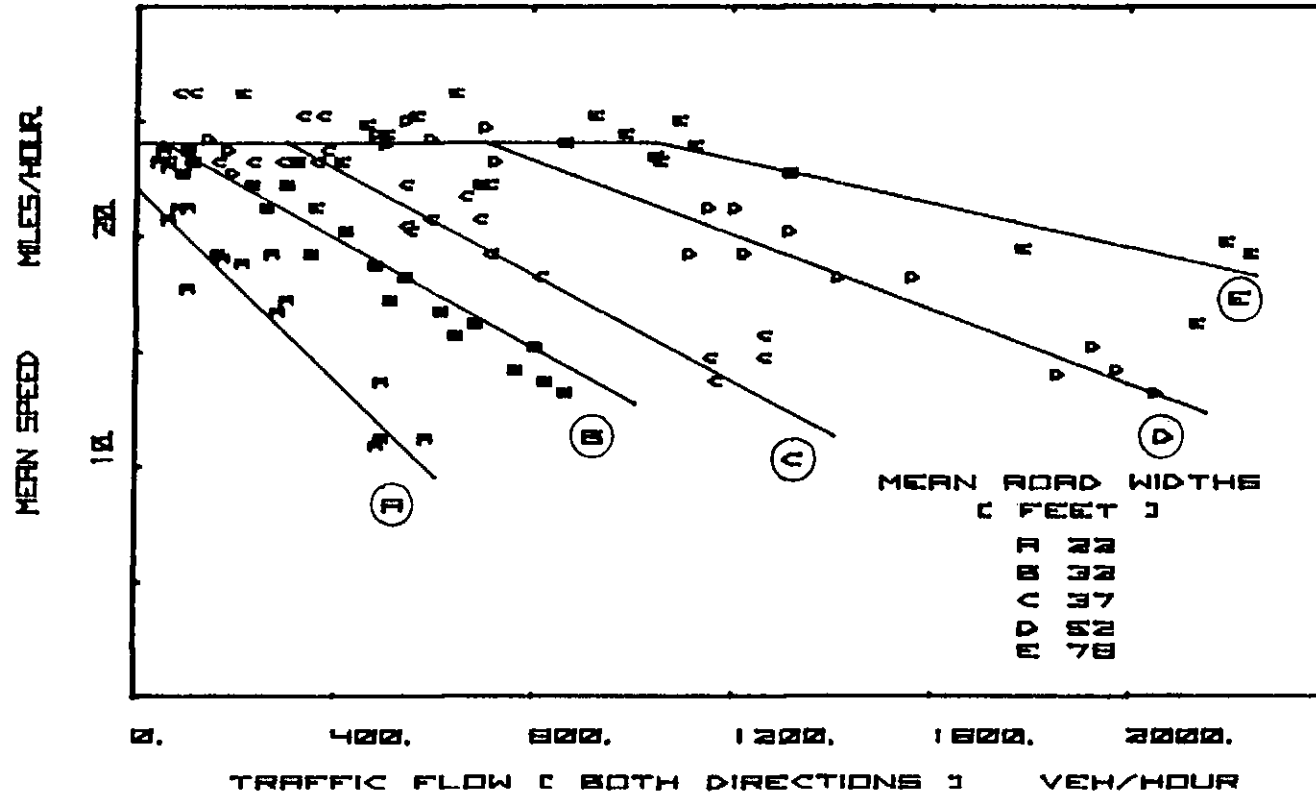


Fig. 2.1 Relation between Mean Speed and Flow in Central London - Wardrop (1952)

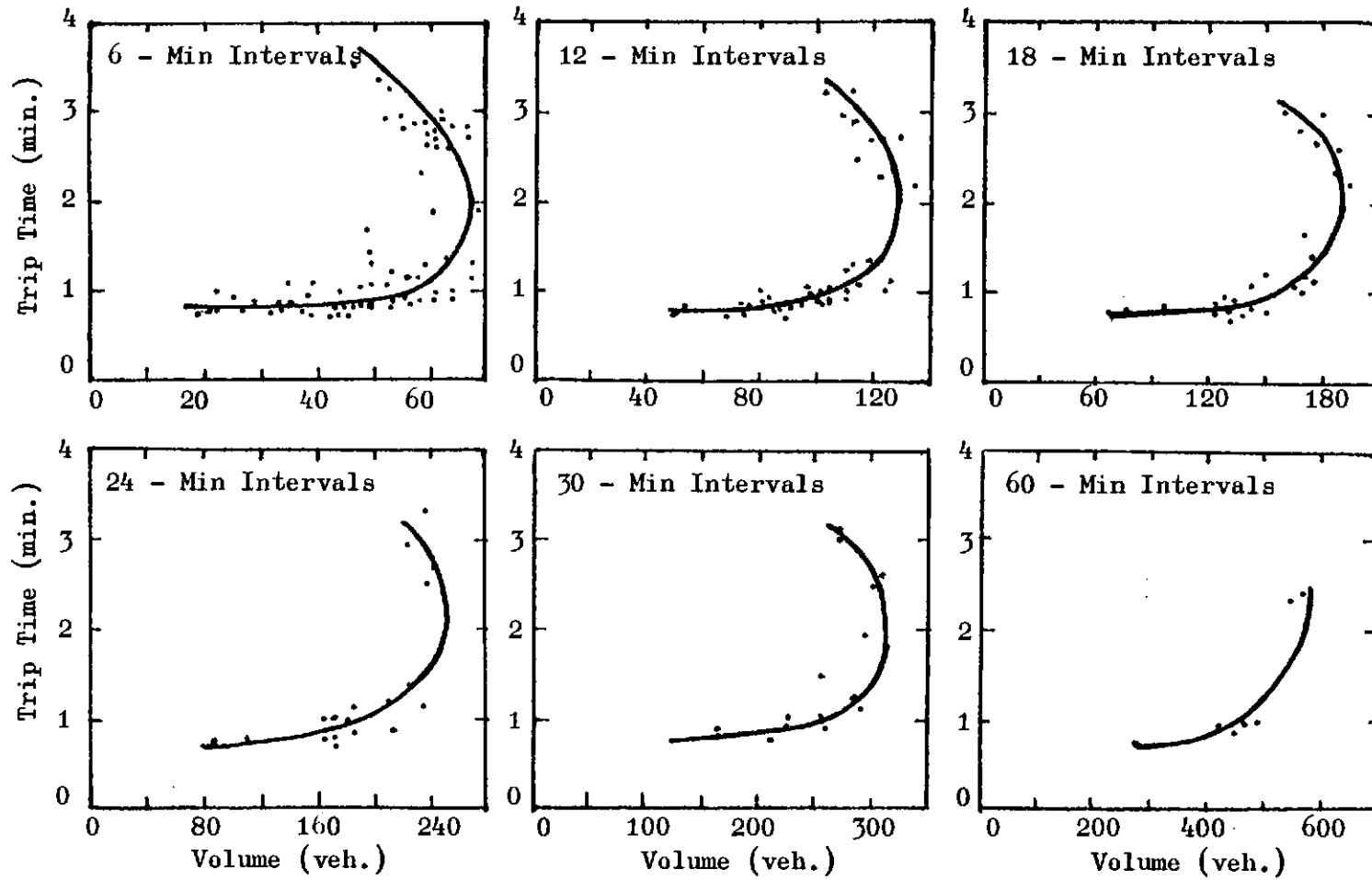
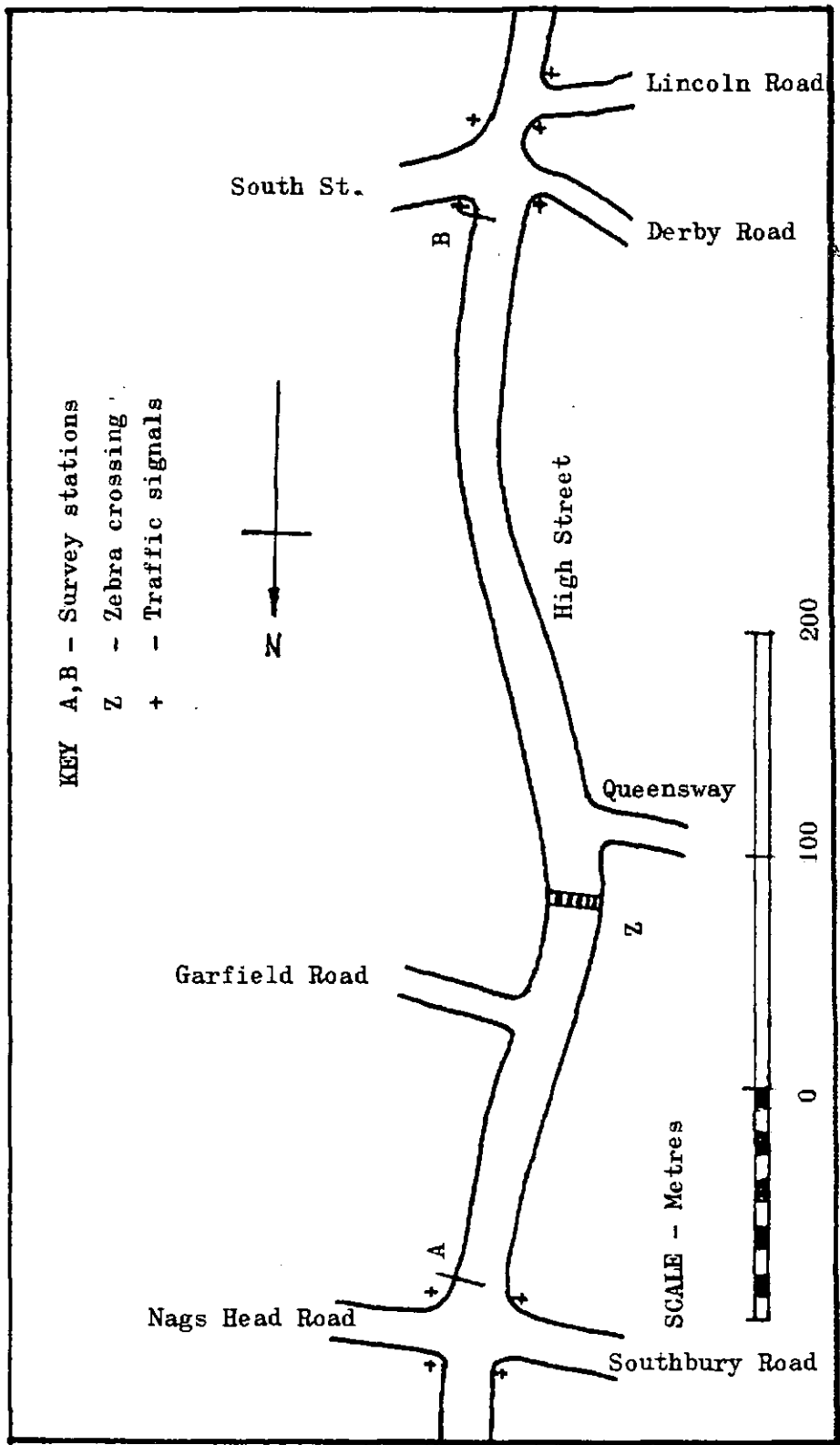
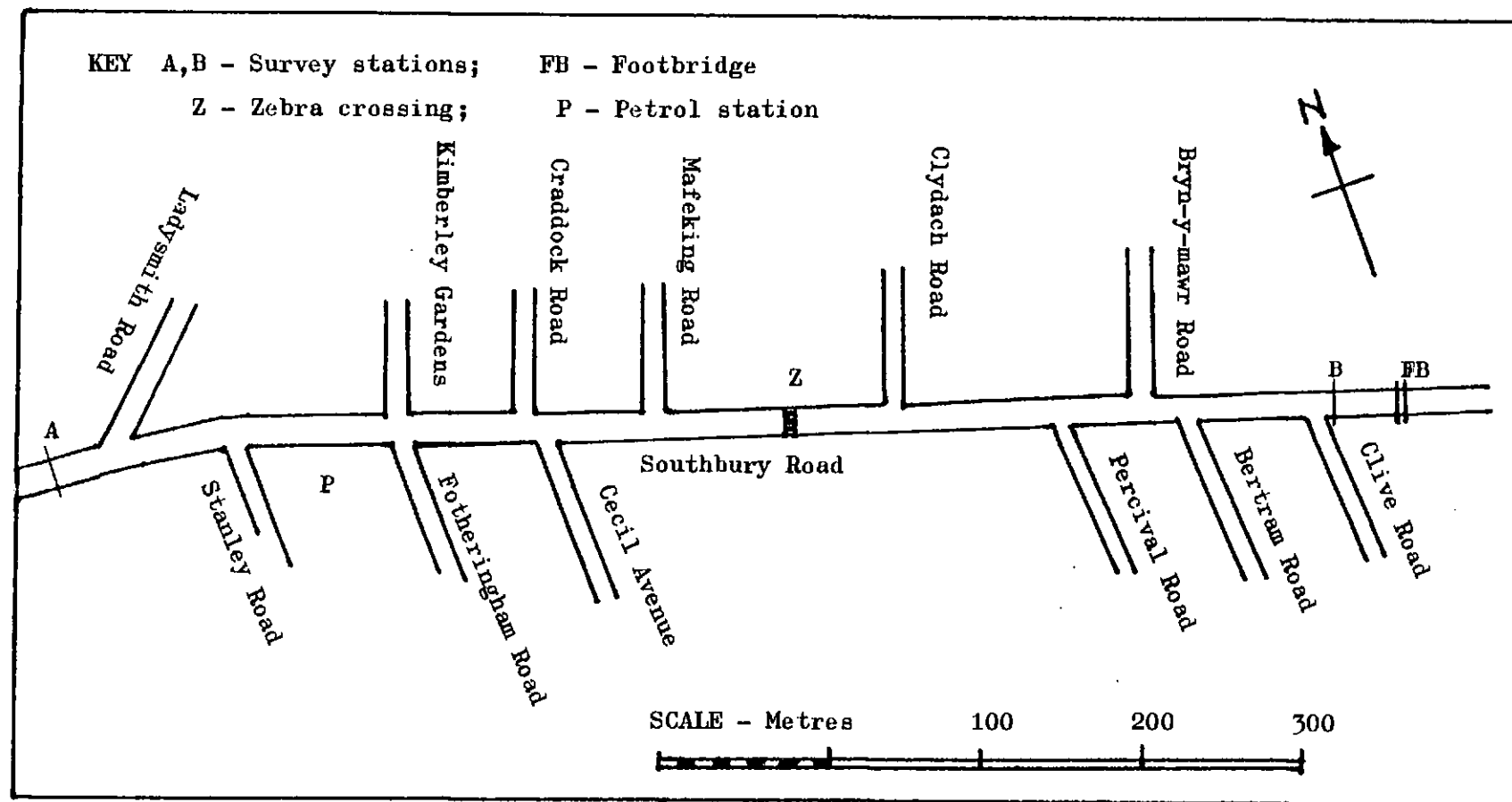


Fig. 2.2 Form of empirical travel time-flow relationships for observation intervals of various durations - Rothrock (1957).



Map 3.1 Location Plan of High Street Study Site.





Map 3.2 Location Plan of Southbury Road Study Site.

Fig. 3.1

HIGH STREET, SOUTHBOUND FLOWS AND JOURNEY-TIMES -  
EVENING PEAK PERIOD, 18/7/72

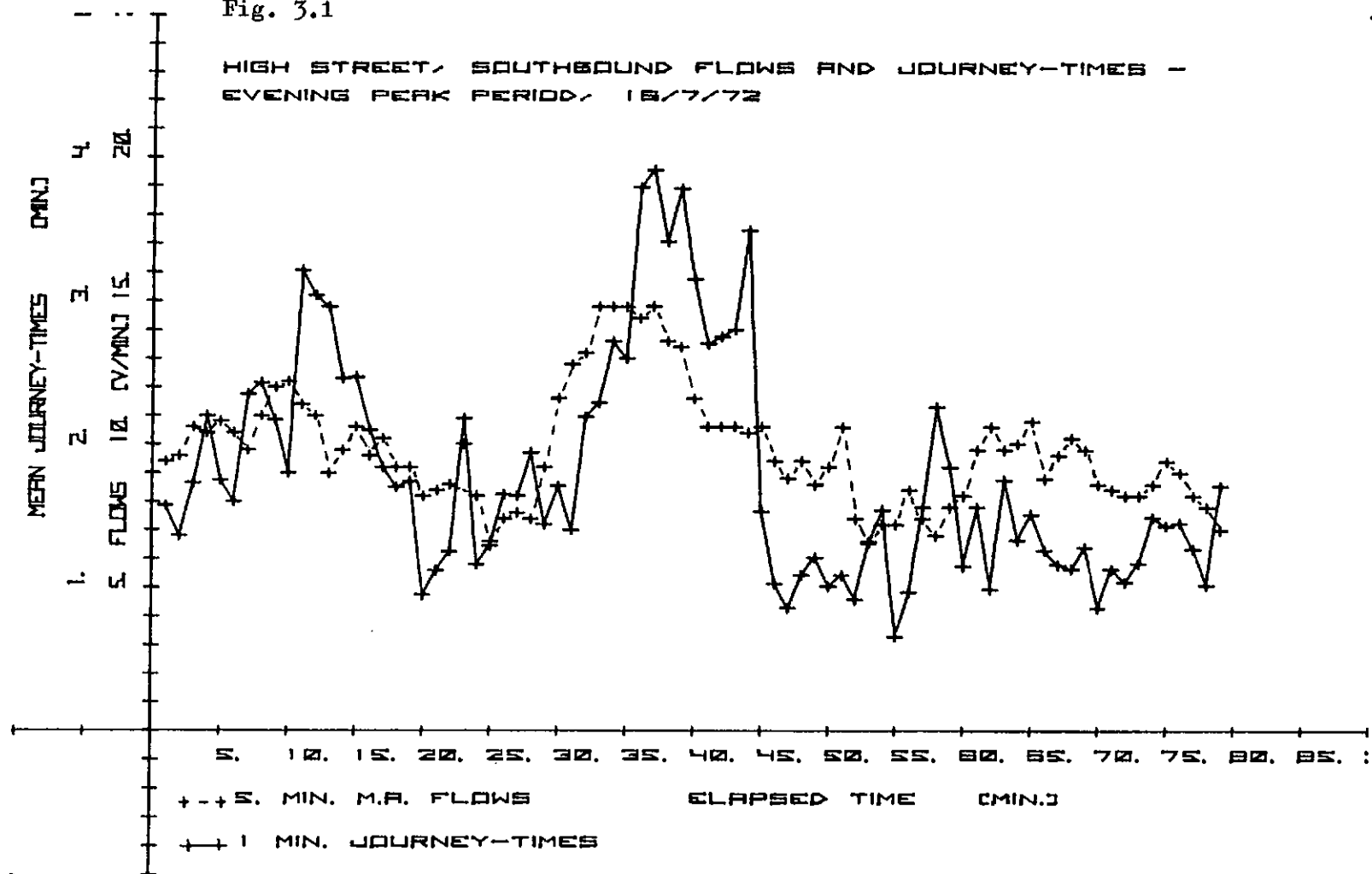
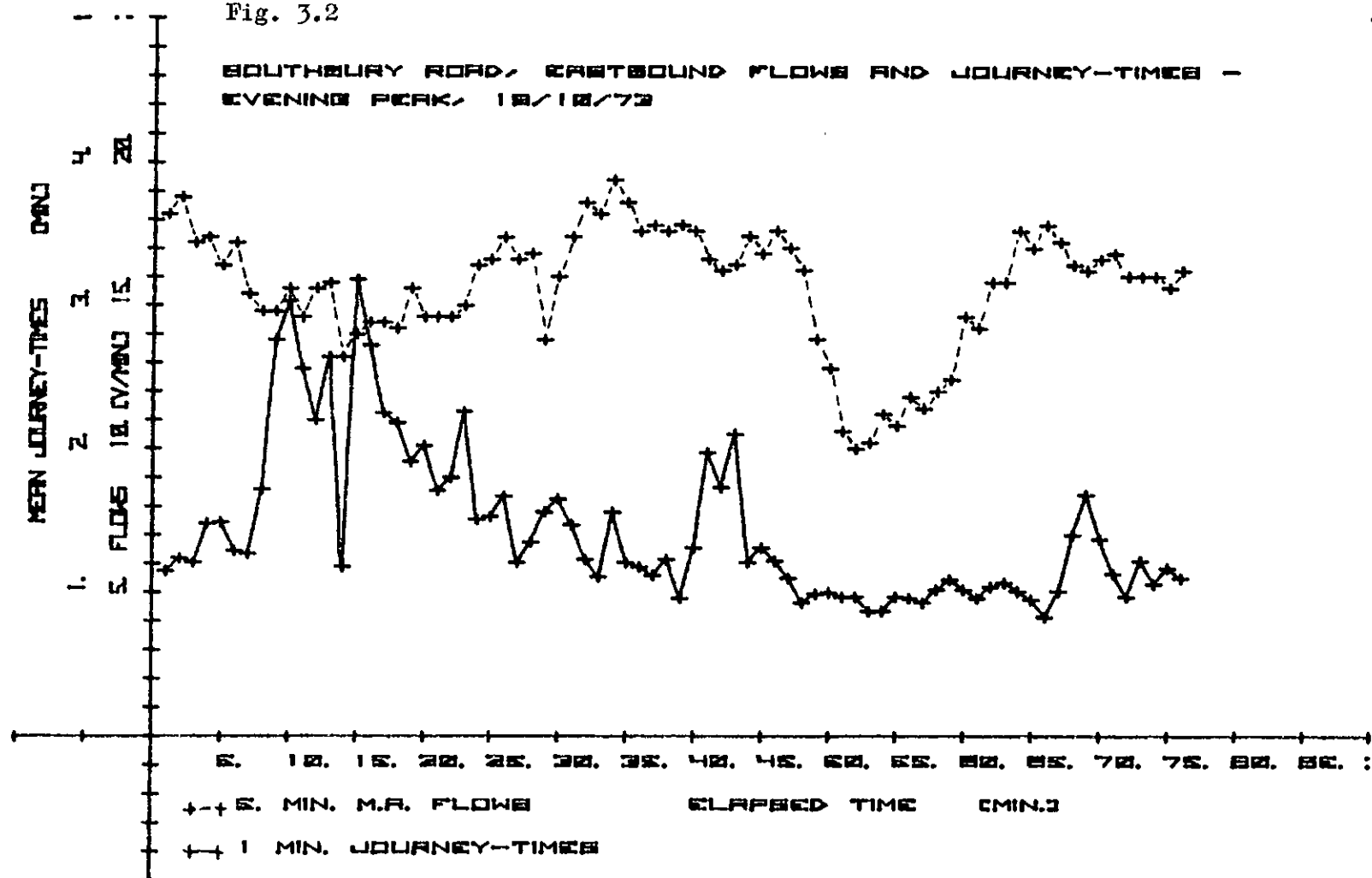
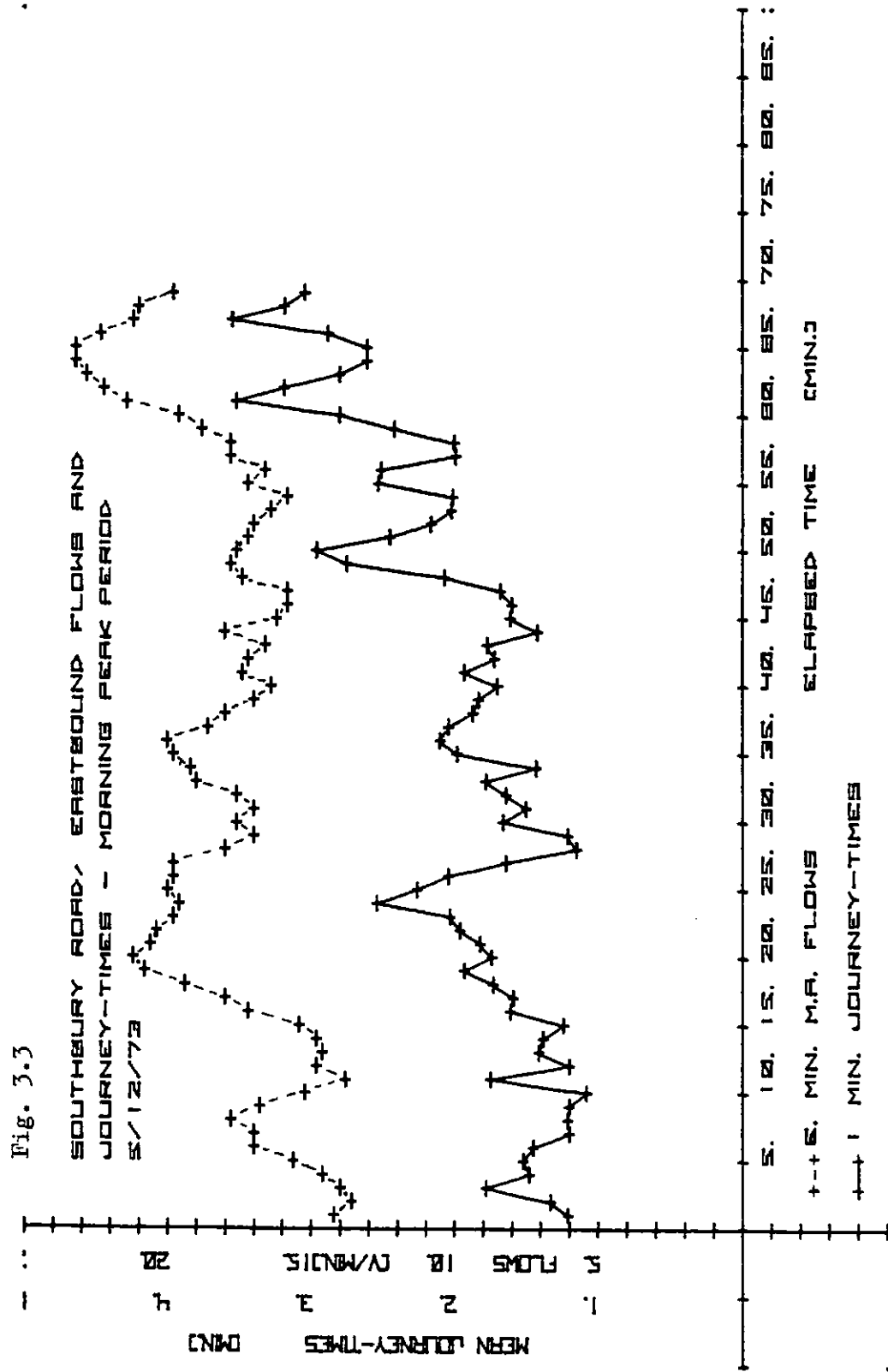
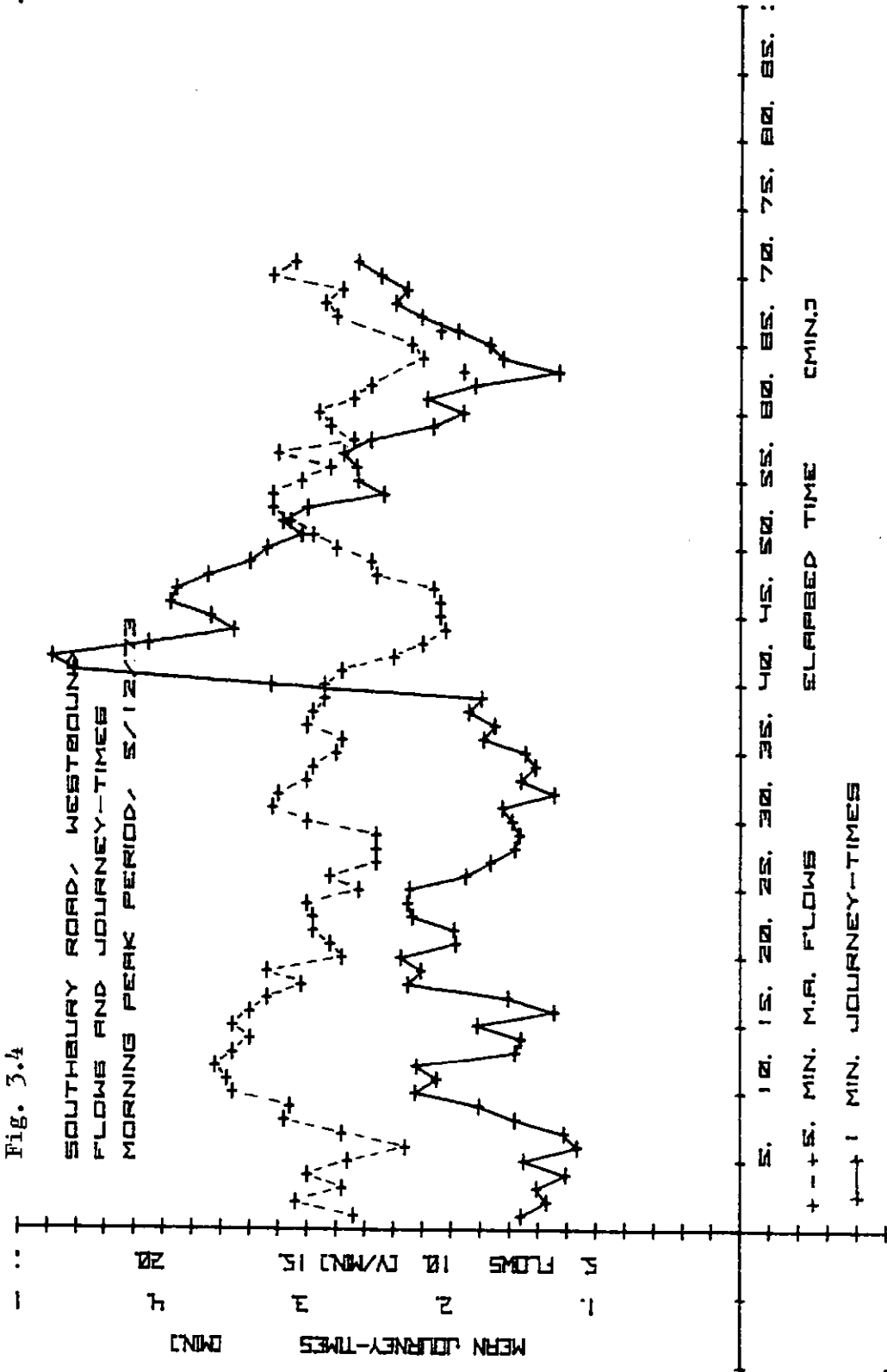


Fig. 3.2

BOUTHURY ROAD, EASTBOUND FLOWS AND JOURNEY-TIMES -  
EVENING PEAK, 18/10/73







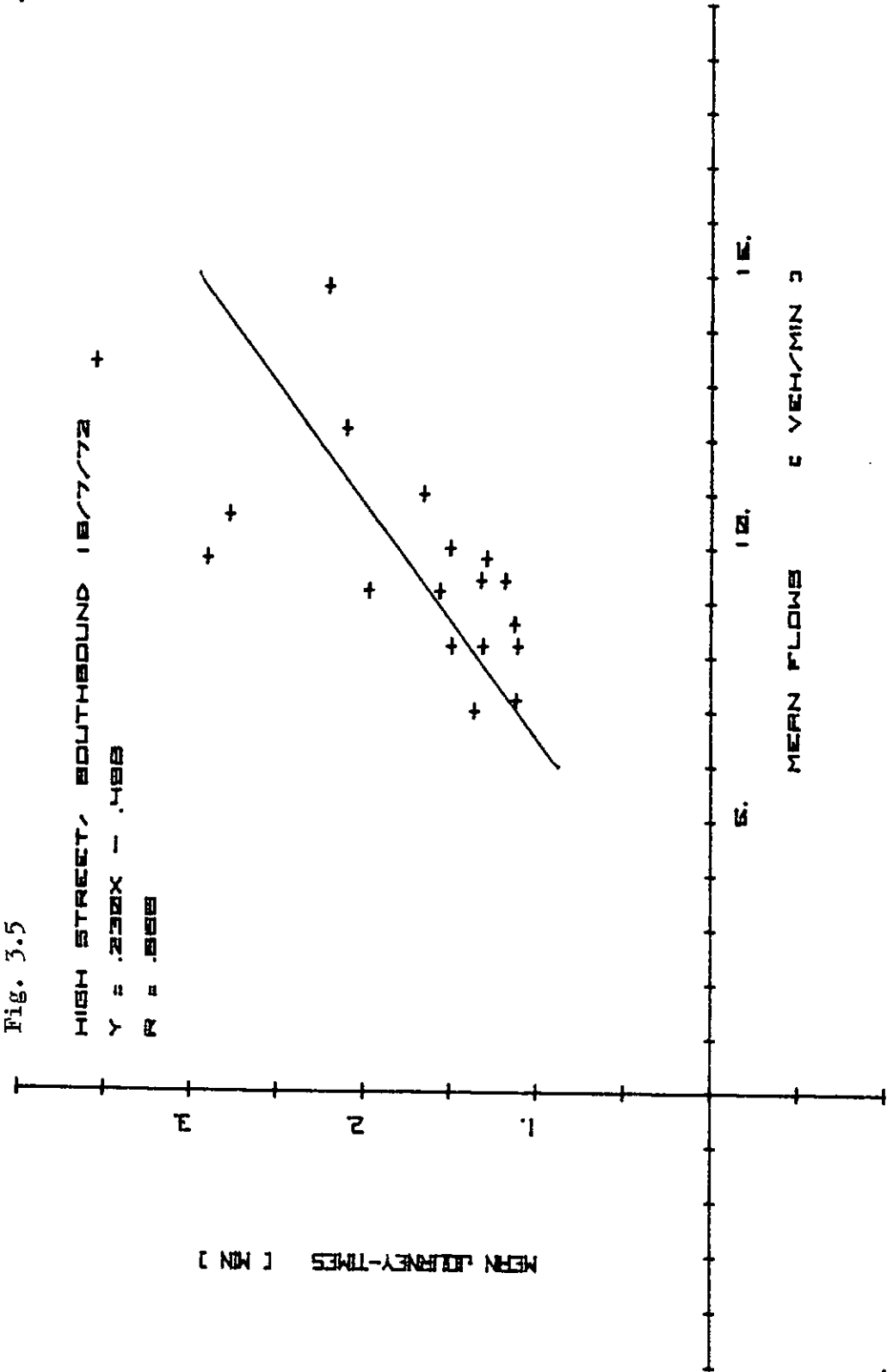
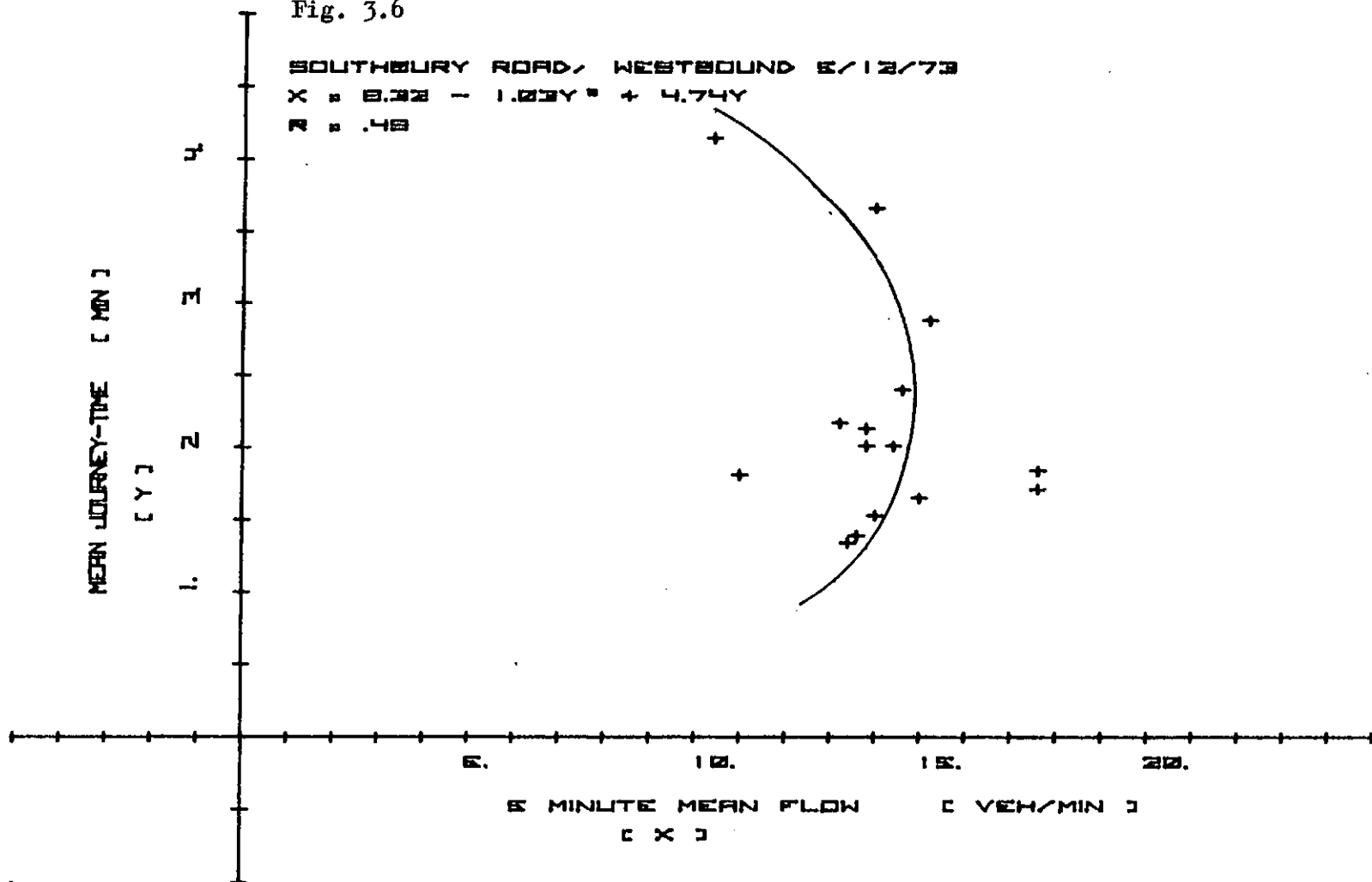


Fig. 3.6

SOUTHBURY ROAD, WESTBOUND 8/12/73

$$X = 8.42 - 1.03Y + 4.74Y$$

$$R = .48$$



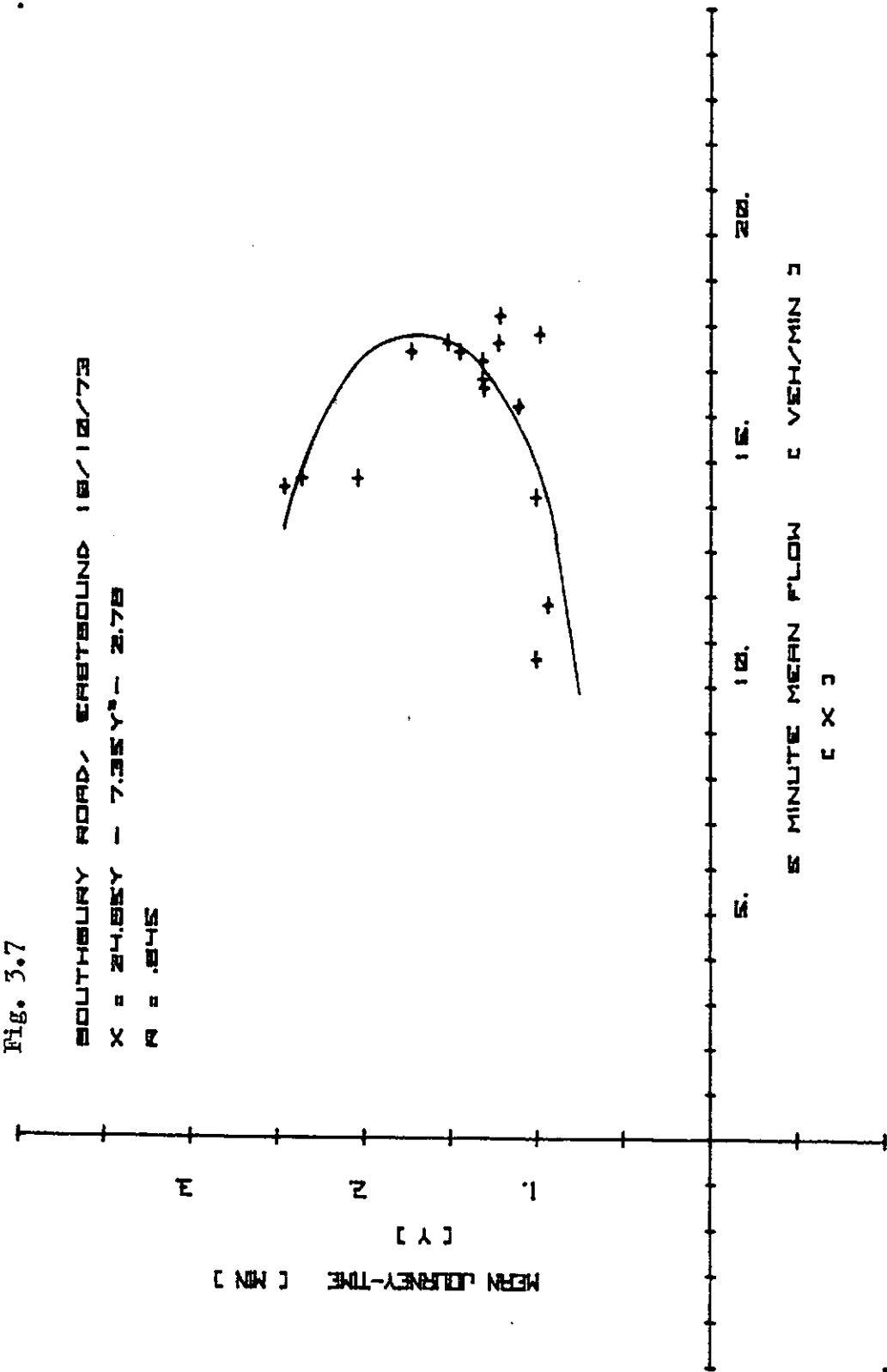
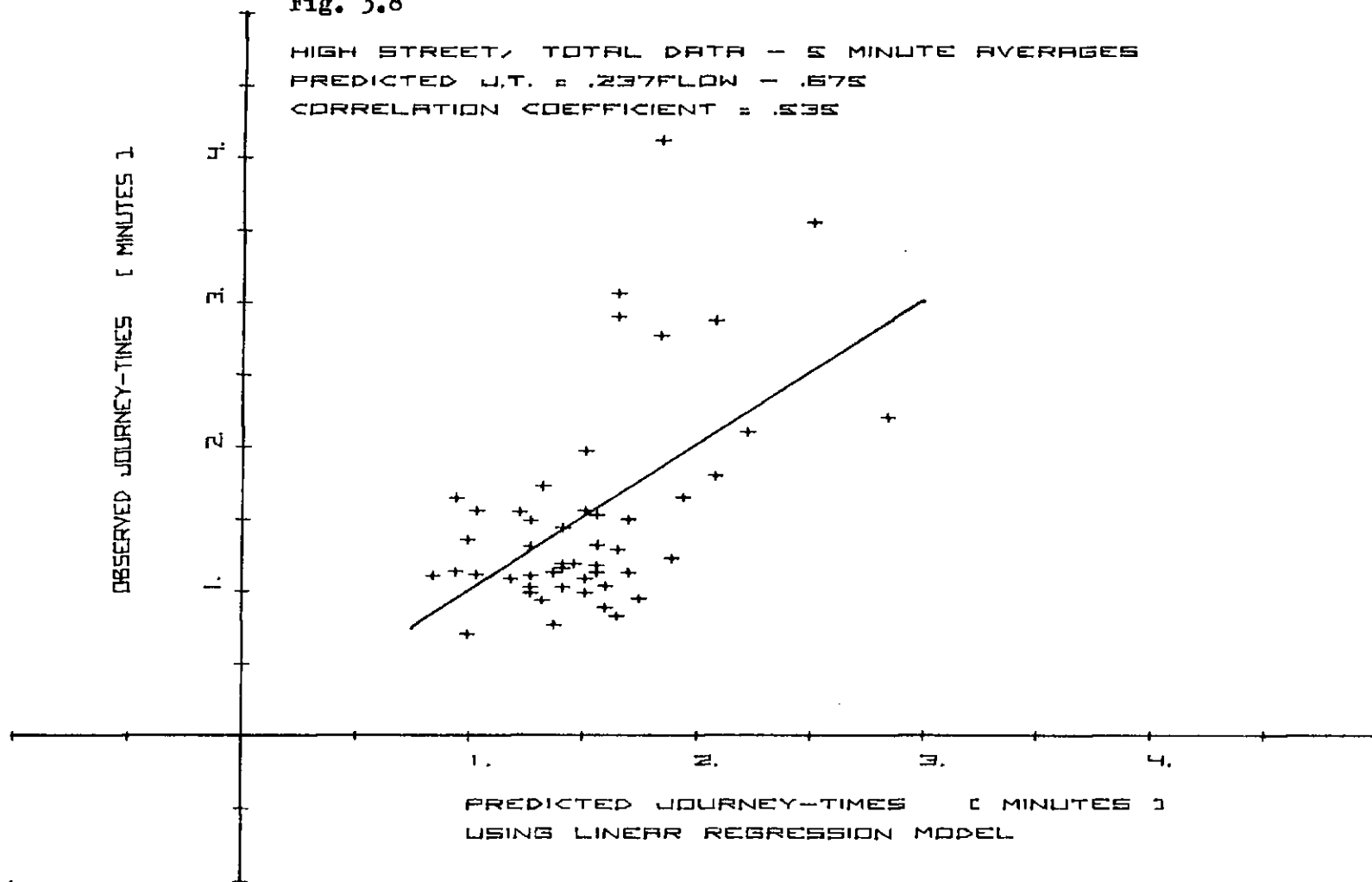
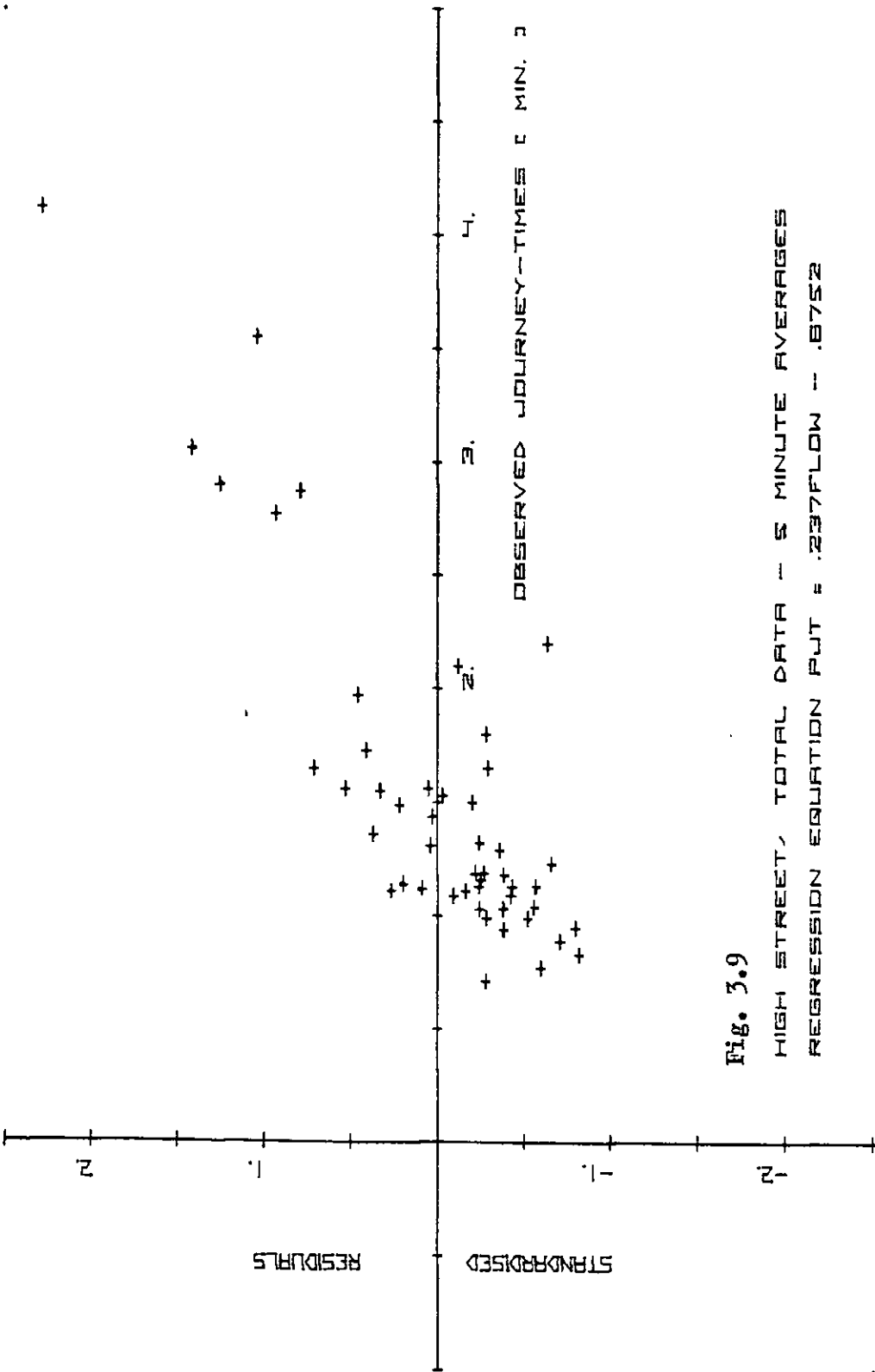




Fig. 3.8

HIGH STREET, TOTAL DATA - 5 MINUTE AVERAGES

PREDICTED J.T. =  $.237\text{FLOW} - .675$ CORRELATION COEFFICIENT =  $.535$ 



# FIGURES

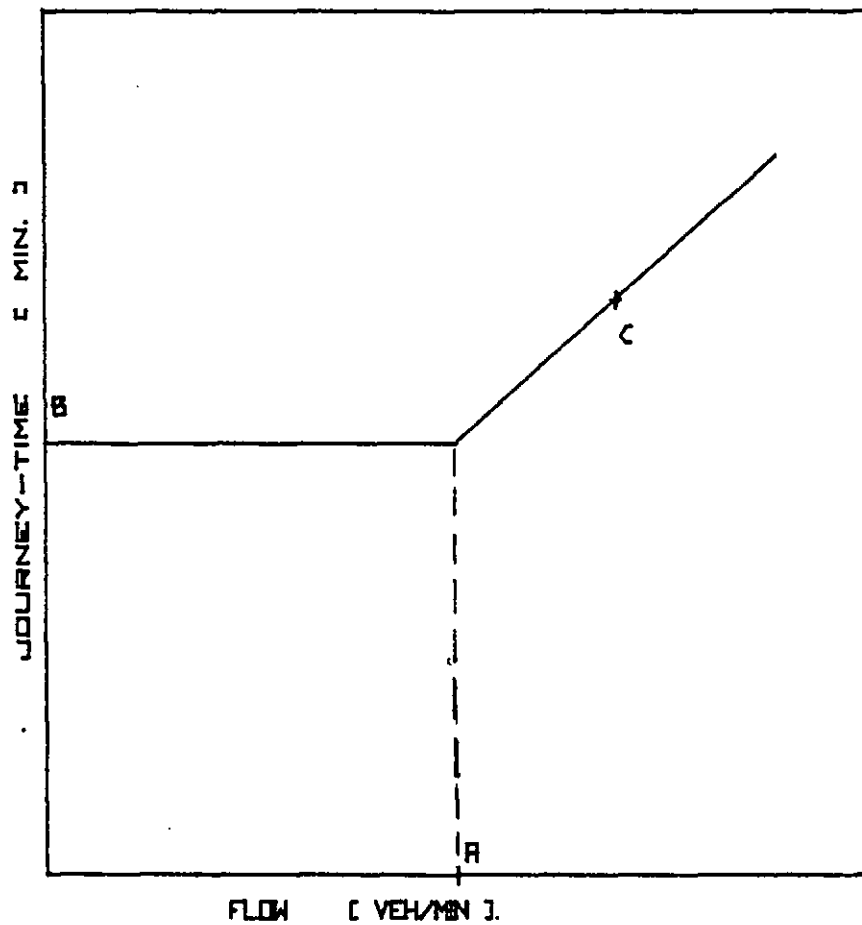
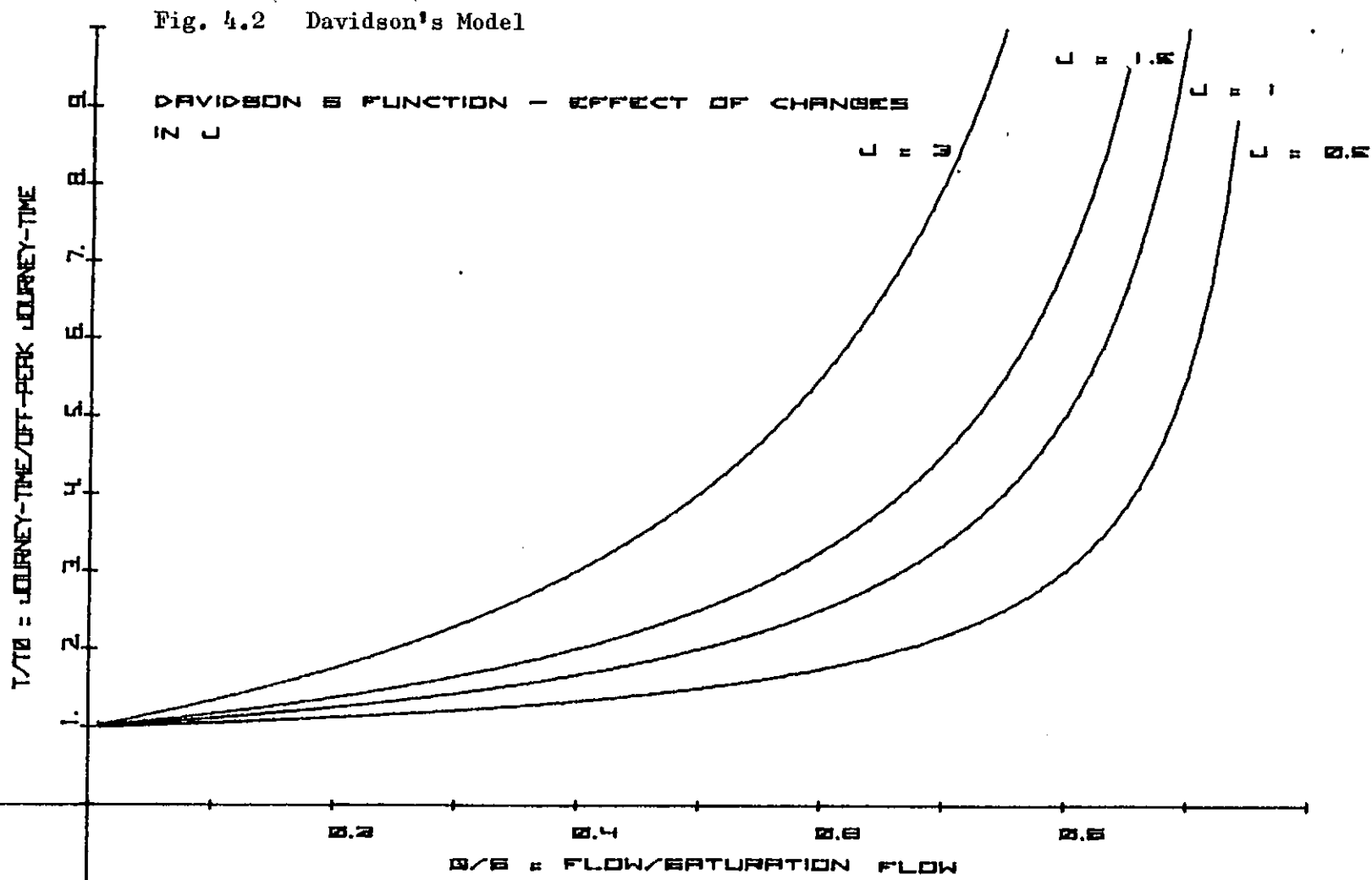


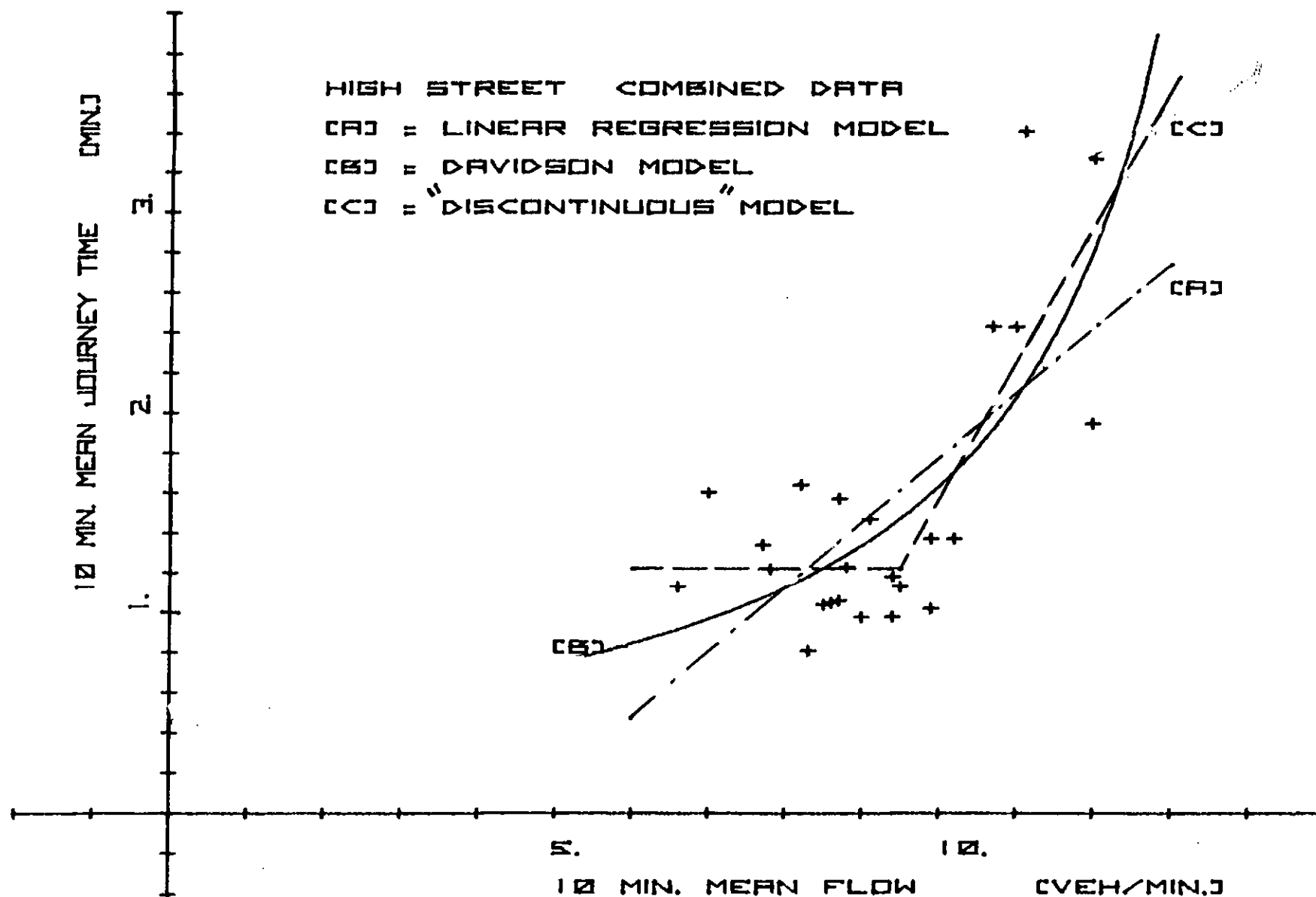
Fig. 4.1 The "Discontinuous" Linear Regression Relationship

$A$  = Assumed critical flow

$B$  = Mean journey-time for flows  $\leq A$

$C$  = Mean flow and journey-time for intervals when flow  $> A$ .





FIGURES

FIG. 4.3. Three Alternative Models of Journey-Time/Flow Relationships

# FIGURES

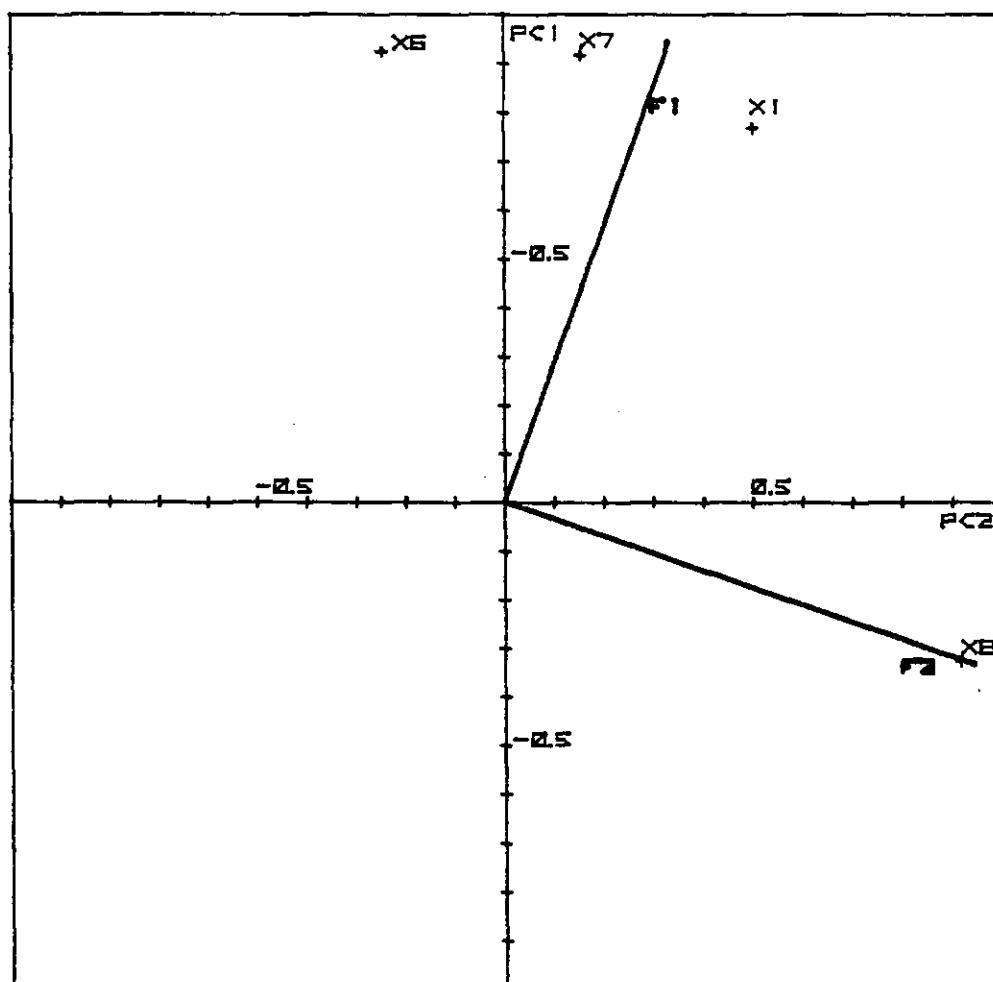


FIG. 5.1 PLOT OF FOUR VARIABLES ON AXES OF THE TWO MOST SIGNIFICANT PRINCIPAL COMPONENTS SHOWING THE AXES AFTER VARIMAX ROTATION - HIGH ST., 18/7/72, 5 MINUTE DATA.

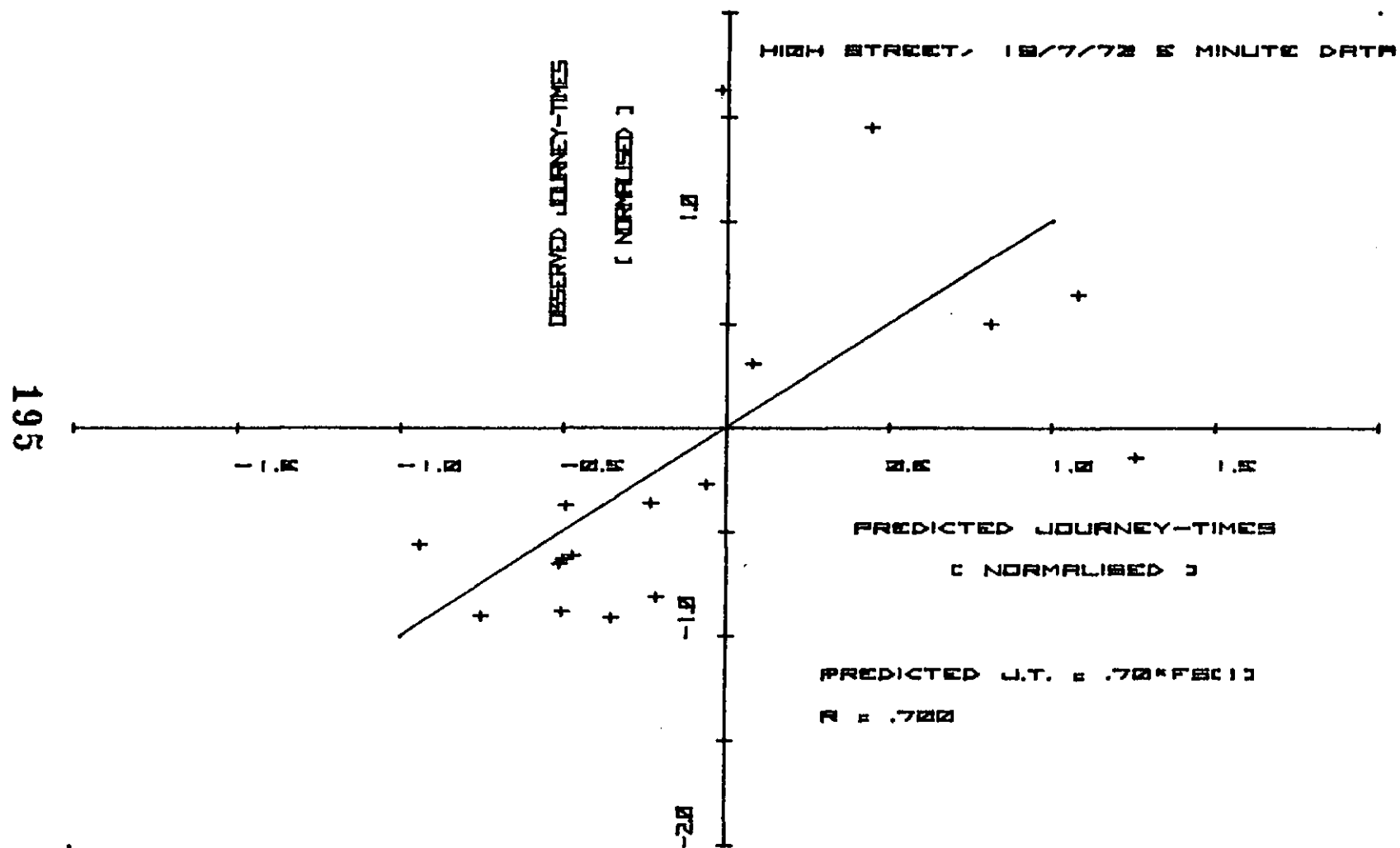


Fig. 5.2 Scatter diagram of Predicted and Observed Journey Times - Factor Model

OBSERVED JOURNEY-TIMES [ MINUTES ]

0.5 1.0 1.5

Fig. 6.1

ENFIELD OFF-PEAK JOURNEY-TIMES  
1978, BASED ON 1978 MODEL  
 $r = .887$

$$\text{PREDICTED J.T.} = 1.10 + .190X_2 - .00240X_1 + .049X_3$$

PREDICTED JOURNEY-TIMES [ MINUTES ]

0.5

1.0

1.5



Fig. 6.2

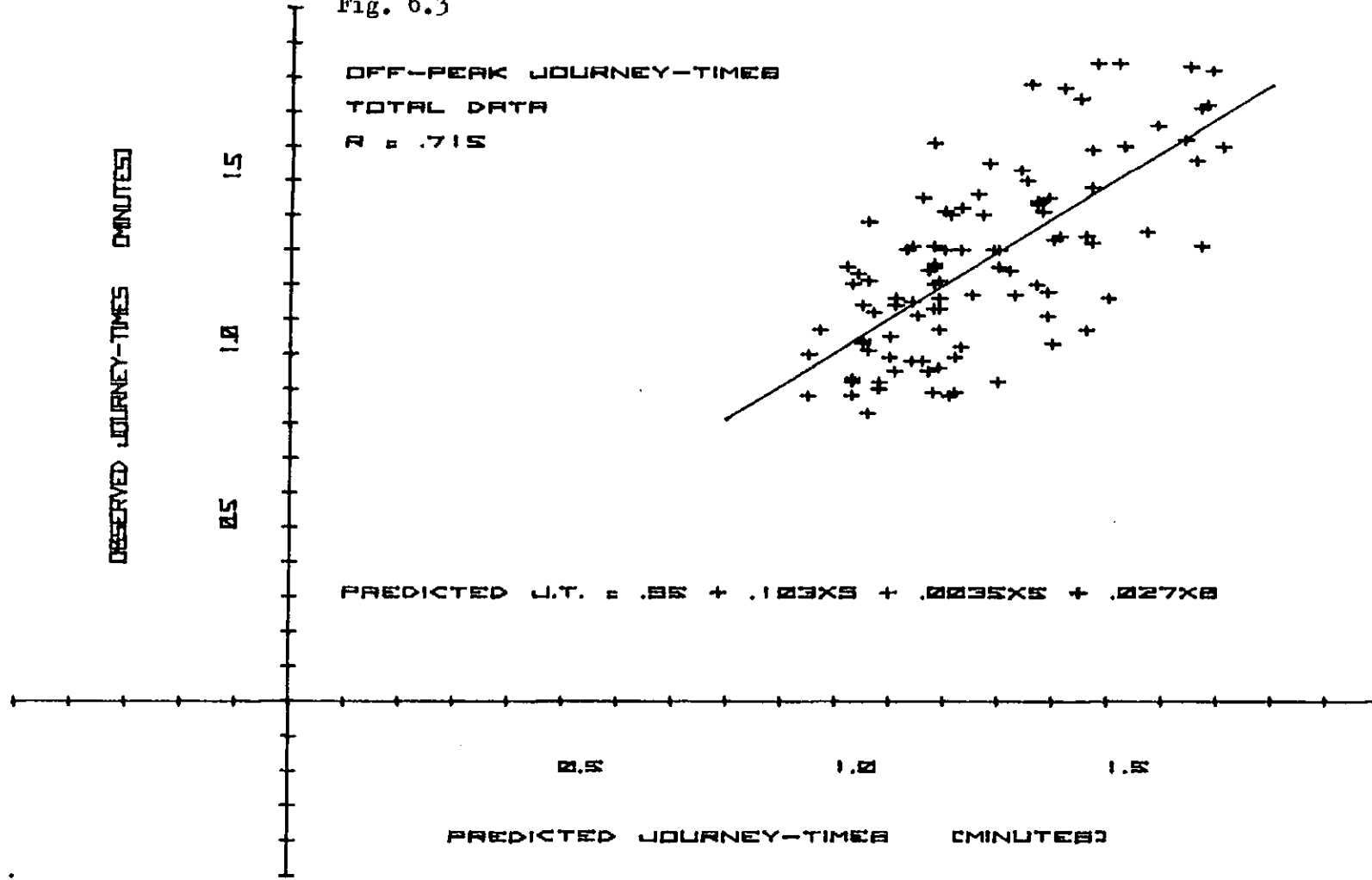
HARINGEY OFF-PEAK JOURNEY-TIMES 1978  
BASED ON ENFIELD (1978) MODEL  
 $R^2 = .801$

$$\text{PREDICTED J.T.} = 1.10 + .138X_2 - .00228X_4 + .043X_8$$

[ SECONDS ] OBSERVED JOURNEY-TIMES

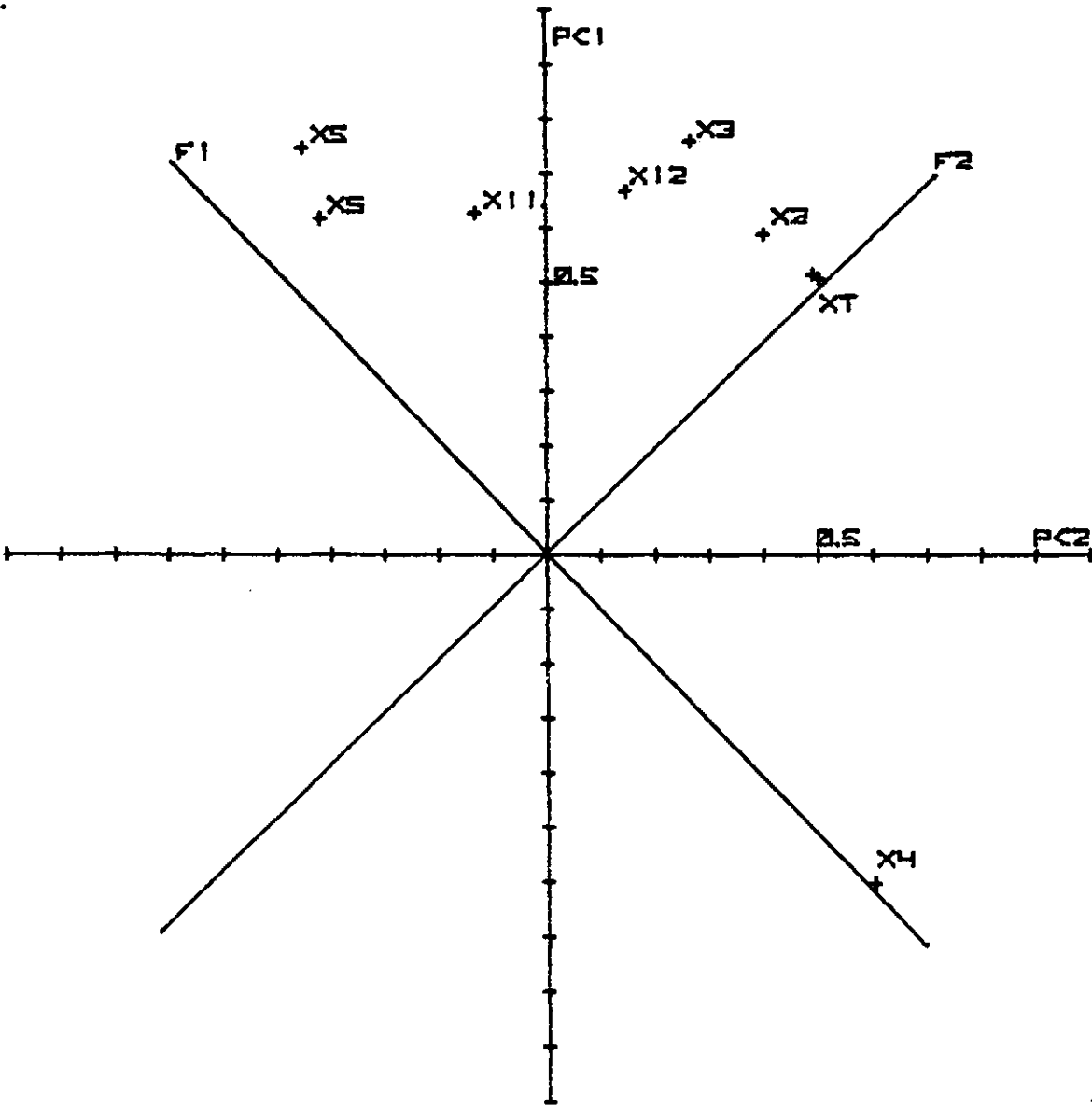
PREDICTED JOURNEY-TIMES [ MINUTES ]

Fig. 6.3



FIGURES

FIG. 7.1 VECTOR DIAGRAM OF FACTOR LOADINGS OF NINE VARIABLES ON TWO FACTORS, SHOWING POSITIONS OF AXES BEFORE AND AFTER ORTHOGONAL ROTATION.



# FIGURES

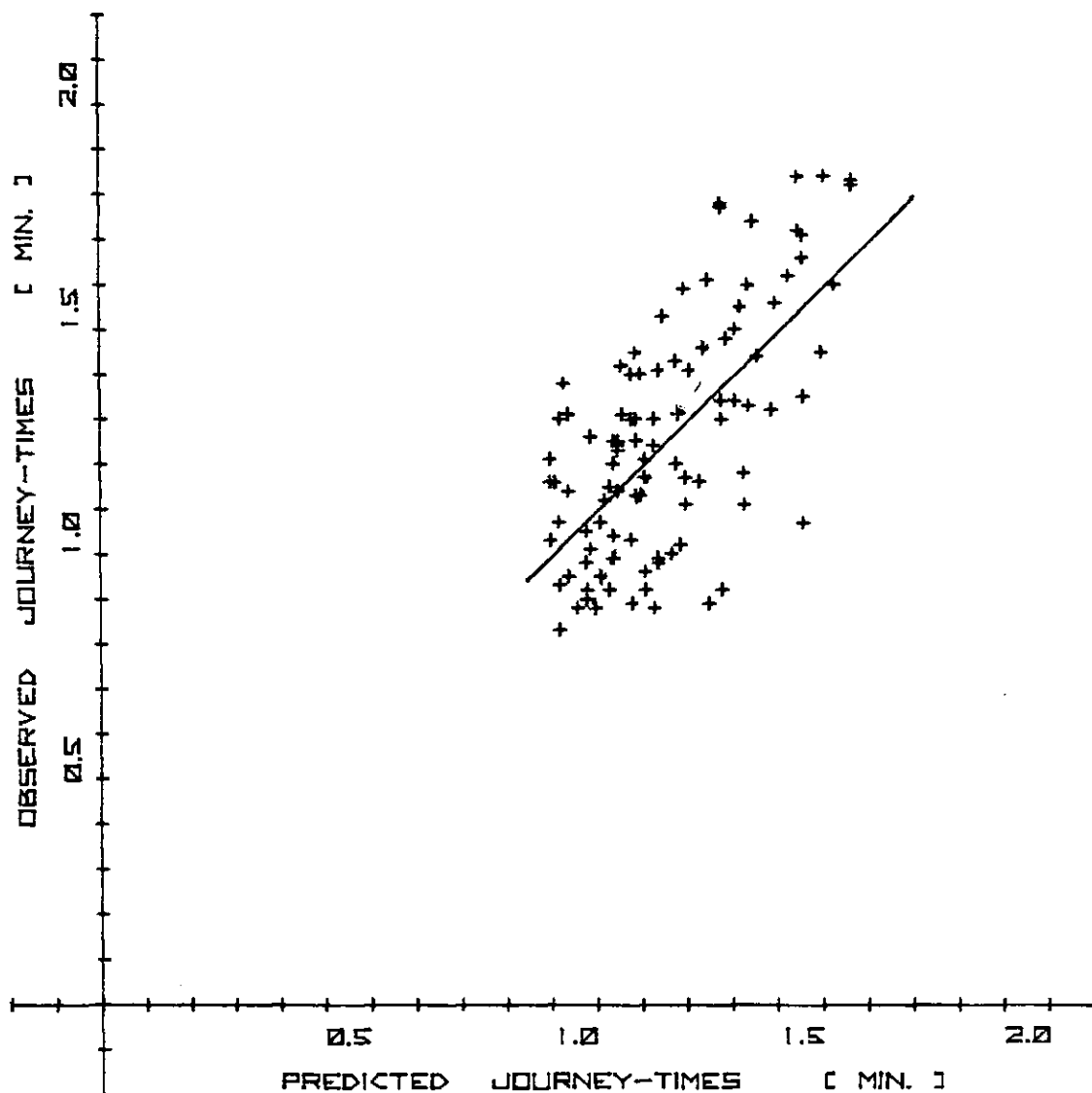


FIG 7.2 SCATTER DIAGRAM OF OBSERVED AND PREDICTED JOURNEY-TIMES, POOLED DATA. PREDICTED JOURNEY-TIMES [ PUT ] GIVEN BY -

$$PUT = 1.26 + .168F1 + .052F2$$

$$R = 0.668$$

## APPENDIX D

### COMPUTER PROGRAMS

Multiple linear regression analysis has been used to a great extent in this study, including a number of transformations of the data where this has been considered to be appropriate. Program FOREG was available as a standard program on the Honeywell computer which was originally used. This is a program which carries out forward stepwise regression and uses selected values of  $F$  in order to test for both inclusion and exclusion of variables at each stage. Various transformations of any or all of the variables are standard options in the program and it is possible to specify which variables may be selected for analysis.

In order to carry out non-standard transformations or aggregation of data special programs were written and a number of these are described below. During the period of this research the Polytechnic purchased a new computer and consequently a period of time was needed to become familiar with the operation of the new system and to alter the necessary programs.

## COMPUTER PROGRAMS

With the new system came a powerful set of statistical programs in a suite called "Statistical Programs for Social Scientists" and these were used for much of the later analyses such as factor and discriminant analysis. The author is grateful to the University of Pittsburgh who have made this suite available.

Program SAMUL was written in order to aggregate the data in any specified number of adjacent sampling intervals and, except for mean journey-time, to express the variables as the average per minute; for example, flow was expressed in vehicles/minute. Hence regression coefficients were always expressed in the same units irrespective of the size of the sampling interval. The use of this program permitted the effects of such aggregations on regression and correlation coefficients to be investigated.

MAMUL was written to carry out a series of aggregations similar to SAMUL but in this case using overlapping moving-averages in order to investigate the trends in the data. The dependent and independent variables can be aggregated over different ranges if desired. In some of the runs only flows were aggregated corresponding to the situation where the variations of the dependent variable were expected to be correlated with variations of the independent variable up to  $K$  intervals earlier, where  $K$  is the number of intervals over which the independent variable was aggregated. The reason for using this definition of moving-average, rather than the usual one of the moving-average representing the value at the midpoint of the range, was because it was considered that journey-times of vehicles were possibly correlated with flows which had occurred earlier but were not correlated with flows which occurred later. In practice some degree

## COMPUTER PROGRAMS

of correlation may exist but this cannot be considered causal in the sense implied in these studies, that changes of flow cause changes of journey-time. A major problem of using overlapping moving-averages is that adjacent values of each variable are correlated with each other after aggregation. The data corresponds to a smoothed time-series rather than random samples from a bivariate distribution. But to some extent this is true of the raw data even if no aggregation is applied, the results of the serial and cross-correlation analysis discussed in Chapter 3 illustrate this. In any case, in this study moving-averages have been used for descriptive purposes only.

Program AUDSR calculates the lagged serial and cross-correlations between flow and journey-times. These calculations were carried out between - flow and lagged flow; flow and lagged journey-time; journey-time and lagged flow; journey-time and lagged journey-time. The calculations were carried out for lags between 0 and 10 minutes.

Program AUQV uses the model derived by Wright(1971) to predict regression slopes and correlation coefficients for long sampling intervals from the lagged serial and cross-covariances between the variables collected in one-minute sampling intervals.

SACOR was written in order to analyse alternative discontinuous regression models of flow and journey-time having a single discontinuity in gradient, see Fig. 4.1. As with program SAMUL the raw data may be aggregated over any desired number of sampling intervals. The program selects a range of alternative points of discontinuity (critical flow) between the observed limits of traffic flow. Below the critical flow variations of journey-time are assumed

## COMPUTER PROGRAMS

to be independent of flow and above the critical flow a linear relationship is assumed. Below the critical flow the mean journey-time corresponds approximately to the mean off-peak journey-time. Above the critical flow the linear relationship is constrained to give the mean off-peak journey-time at the critical flow and to pass through the mean values of these variables above the critical flow. To permit comparisons to be made between the results of choosing different positions for the point of discontinuity the residual sum of squares and the mean squares are determined. The same statistics can be used to compare the results with those given by conventional linear regression analysis.

Program LLCOR carries out a discontinuous analysis using a power curve to define the relationship above the critical flow and to have zero slope at the critical flow.

Another program which carries out analysis similar to the last two programs is program DAJ. This program investigates the use of Davidson's model for predicting the variation of journey-time with flow. Initial values of off-peak journey-time are entered as well as the usual data file, the program calculates the approximate J factor and the mean square error of the predicted journey-times. The values of off-peak journey-times and saturation flows are altered incrementally and new values of J and mean square error are calculated. One hundred pairs of values of saturation flows and off-peak journey-times are used and the output is printed in a table. The user inspects the table to determine the optimum value of J corresponding to the minimum mean square error.



## COMPUTER PROGRAMS

A number of other shorter programs have been written to carry out some basic calculations such as summing variables and their squares, calculating regression and correlation coefficients, and calculating predicted values of the dependent variables. These short programs were sometimes interconnected with larger ones where the processes were to be repeated. All programs were written to accept data in the same standard format in order to avoid restructuring data before analysis.

Finally, the text of this thesis was produced by a document-editing program called RUNOFF and the graphs were produced on a graph plotter controlled by a Hewlett-Packard 9810 programmable calculator.

APPENDIX E  
PRINCIPLES OF FACTOR ANALYSIS

In Chapter 5 the similarities between two multivariate techniques of analysis were discussed, these techniques were Principal Component Analysis and Common Factor Analysis and in Chapters 5 and 7 applications of the techniques were considered. In this appendix their theoretical principles will be explained.

E.1 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis attempts to simplify the data by finding a number of new variables, called principal components, which are uncorrelated and mutually orthogonal linear combinations of the original variables. Supposing that observations of  $p$  variables  $x$  have been made on  $N$  individuals. The  $p$  new variables  $e$  will be of the following form:

$$e_i = \sum_{k=1}^p l_{ik} x_{ik}$$

Almost invariably, the analysis first normalises the  $x$ 's before continuing with the remainder of the process, using the following

## PRINCIPLES OF FACTOR ANALYSIS

relationship:

$$z_i = \frac{x_i - \bar{x}_i}{Sx_i}$$

where  $z_i$  = normalised value of variable  $x_i$

$\bar{x}_i$  = mean value of  $x_i$

$Sx_i$  = standard deviation of  $x_i$ .

For the  $e$ 's to be uncorrelated then,

$$\text{Exp } (e_i e_j) = \text{Exp } \sum_{k=1}^p \ell_{ik} \cdot z_k \sum_{m=1}^p \ell_{jm} z_m$$

Hence:

$$\sum_{k,m=1}^p \ell_{ik} \cdot \ell_{jm} \cdot \text{Exp } (z_k z_m) = 0$$

or

$$\sum_{k,m=1}^p \ell_{ik} \cdot \ell_{jm} \cdot r_{km} = 0$$

where  $r_{km}$  = correlation between  $x_k$  and  $x_m$

since the  $x$ 's are normalised. Assuming no  $r_{km}$  is zero, which is usually the case, then we are imposing the condition that

$$\sum_{k,m=1}^p \ell_{ik} \cdot \ell_{jm} = 0 \quad \text{for } i \neq j$$

Since we require the components to be mutually orthogonal then

$$\sum_{k,m=1}^p \ell_{ik} \cdot \ell_{im} = 0 \quad \text{for } k \neq m \quad (1)$$

$$\sum_{k=1}^p \ell_{ik}^2 = 1 \quad \text{for } k = m \quad (2)$$

## PRINCIPLES OF FACTOR ANALYSIS

Matrix notation is more convenient for developing the basic theory.

Equation (1) can be represented thus:

$$E = LZ$$

where  $Z$  is a column vector representing the  $Z$ 's

$E$  is a column vector representing the  $e$ 's

$L$  is a  $(p \times p)$  matrix of constants.

Premultiplying (2) by  $L'$ , the transpose of  $L$ , gives

$$L'E = L'LZ$$

but from (1)

$$L'L = I,$$

where  $I$  is the identity matrix, and so

$$Z = L'E$$

Now considering the covariances of the  $e$ 's,

$$\begin{aligned} \text{Exp}(EE') &= \text{Exp}(LZ(LZ)') \\ &= \text{Exp}(LZZ'L') \end{aligned}$$

Putting

$$\Delta = LZZ'L'$$

$\Delta$  is a matrix of the following type:

$$\Delta = \begin{bmatrix} \lambda_1 & 0 & 0 & . & . \\ 0 & \lambda_2 & . & . & . \\ 0 & . & . & . & . \\ 0 & . & . & . & . \\ . & . & . & . & \lambda_p \end{bmatrix}$$

## PRINCIPLES OF FACTOR ANALYSIS

since the  $e$ 's are uncorrelated with each other. Now premultiplying by  $L'$  gives

$$XX'L' = L'\Delta$$

This can be rewritten as

$$(R - I)L' = 0$$

Thus the values of  $\lambda_1, \lambda_2$  etc are the eigenvalues derived by the solution of (3).

The functions  $e_1, e_2$  etc are the principal components and these are usually listed in descending order of eigenvalues since it can be shown that the proportion of variance in the original data which is accounted for by each principal component is proportional to the corresponding eigenvalue.

### E.2 COMMON FACTOR ANALYSIS

The common factor model can be represented as follows:

$$Z_{ij} = C_{j1} h_{i1} + C_{j2} h_{i2} + \dots + C_{jp} h_{ip} + d_j U_{ij}$$

or

$$Z_{ij} = \sum_{k=1}^p C_{jk} h_{ik} + d_j U_{ij}$$

where  $z_{ij}$  = the normalised value of the  $j$ th variable  
in the  $i$ th sample,

$h_{ik}$  = the  $i$ th value of the  $k$ th common factor,

$u_{ij}$  = the  $i$ th value of the unique (non-shared)  
factor associated with variable  $j$ ,

$c_{jk}$  = the weight assigned to the dependence of  
the  $j$ th variable on the  $k$ th common factor,

## PRINCIPLES OF FACTOR ANALYSIS

$d_j$  = the weight assigned to the  $j$ th unique factor.

Thus the weights are similar to regression coefficients and determine the observed values of the variables.

The model appears to be much less parsimonious than the principal component model since it involves  $p$  common factors and  $n$  unique factors. However, the unique factors are solely intended to permit the unique variance to be eliminated from the relationships between the original variables and the analysis concentrates on the common factor portion of the variables. The model represents the common proportion of each variable by the factors  $h_j$  and the unique portion by the unique variable  $u_j$ . Thus each variable may be regarded as having an element particular to itself and separate from the common causal phenomena; alternatively the residuals may be regarded as unique effects of the measured variables in contrast to the common effects of the common factors. The proportions of the variation of each variable attributable to each part are called uniqueness and communality respectively.

In common factor analysis the techniques attempt to fit  $p$  (less than  $n$ ) factors to a set of  $n$  variables, such that the observed covariance matrix can be reproduced by

- (a) a  $p$  dimensional covariance structure (derived from the common factors); and
- (b) a set of  $n$  error variances (i.e. a diagonal covariance matrix).

## PRINCIPLES OF FACTOR ANALYSIS

Since the variance-covariance matrices are derived from normalised data the solutions are not affected by the varying scales of the original data. Also the derivation of a diagonal covariance matrix, mentioned in (b), implies that the unique variables are mutually independent.

In matrix notation the common factor model can be represented as follows:

$$Z = C.H + D.U$$

where  $Z$  = matrix of normalised variables ( $n \times N$ )

$C$  = matrix of common factor weights ( $n \times p$ )

$H$  = matrix of values of common factors ( $p \times N$ )

$D$  = diagonal matrix of unique factor weights ( $n \times n$ )

$U$  = matrix of values of unique factors ( $n \times N$ )

The term  $C.H$  can be considered as an approximation to the original data in a lower dimensional form. The common factors may be intercorrelated or orthogonal and the analyst is at liberty to choose which form of analysis is most appropriate on "a priori" grounds.

Since the variables have been normalised the covariance matrix is identical to the correlation matrix. This matrix can be partitioned in a manner which reflects the underlying structure of the common factor model, thus

$$R = R_r + D' D$$

## PRINCIPLES OF FACTOR ANALYSIS

$$= \begin{bmatrix} h_1^2 & r_{12} & . & . & r_{1n} \\ r_{12} & h_2^2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ r_{1n} & . & . & . & h_n^2 \end{bmatrix} + \begin{bmatrix} d_1^2 & 0 & . & . & . \\ 0 & d_2^2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & d_n^2 \end{bmatrix}$$

The terms in the diagonal of  $R$  are the communalities of the measured variables, these being related to the common factor model as follows;

$$h_j^2 = c_{j1}^2 + c_{j2}^2 + \dots + c_{jn}^2$$

The term  $d$  is known as the uniqueness of the variable  $j$ . Thus the uniqueness and communality of a variable are related by the following:

$$h_j^2 + d_j^2 = 1$$

It can also be shown that the matrix  $R$  can be written in the form:

$$R_r = C R_h C'$$

where  $R_h$  is the inter-factor correlation matrix. For an orthogonal factor structure this is an identity matrix. Thus the analytical problem of common factor analysis is to determine the factor pattern matrix which will reproduce the reduced correlation matrix  $R_r$ . However, there is one great difficulty and that is the communalities in the major diagonal of  $R_r$  are unknown. For less than four measured variables theoretical values of the communalities can be deduced, but for greater numbers of variables estimated values are inserted and an iterative procedure adopted in order to determine the factor pattern matrix. After each iteration, the derived factor pattern ( $c_{j1}, c_{j2}$  etc) is used to recalculate the communalities and a new factor pattern is determined. The process is then repeated until minimal changes occur in the values



## PRINCIPLES OF FACTOR ANALYSIS

of the communalities between successive factor patterns.

The matrix  $R_r$  can also be represented by:

$$R_r = S.C'$$

where  $S = C.R_h$

$C'$  = the transposed matrix  $C$

The matrix  $S$  is known as the factor structure matrix and this indicates the correlation between the factors and the measured variables. If the factors are orthogonal the  $S$  and  $C$  matrices are identical. Usually factor analysis produces matrices of factor loadings which are complex, with most of the variables being loaded on (correlated with) a large number of factors and making the factors difficult to interpret. However, the pattern of factors which account for the inter-correlations is not unique and it is possible to carry out linear transformations of the factors in order to produce a factor pattern which is easier to interpret. This process is known as rotation, and is intended to produce factor matrices with each variable loaded heavily on a few factors and insignificantly loaded on the remainder. This is the condition which Thurstone (1965) calls "simple structure". Thus an individual factor can be associated with the particular group of variables on which it is most heavily loaded and it may be possible to identify the underlying nature of this association.

### E.3 ROTATION OF AXES.

A number of rotation procedures have been devised, such as Varimax by Kaiser(1958) and Quartimax by Carroll(1953); these are orthogonal rotation methods. In addition, oblique factor rotation can be carried out wherein it is assumed that some intercorrelation between factors is likely, but in this case the analysis is more complicated. The most widely used method is Kaiser's method and this is based on the principle that factor patterns which are most easily interpreted have either high or low factor loadings, and no intermediate ones. Such factor patterns have a high variance of the squared factor loadings and thus the criterion adopted is to maximise this variance.

The process of rotation can be most easily understood by using a diagram to demonstrate the method for a two-factor model.

If there are  $n$  variables then the initial factor pattern can be represented by a factor loading matrix  $C1(n \times 2)$  and the rotated factor pattern by the matrix  $C2(n \times 2)$ . Thus:

$$C1 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ . & . \\ x_{n1} & x_{n2} \end{bmatrix}$$

and:

$$C2 = \begin{bmatrix} x_{11}' & x_{12}' \\ x_{21}' & x_{22}' \\ . & . \\ x_{n1}' & x_{n2}' \end{bmatrix}$$

In Fig E.1 the points 1,2,3 ...  $n$  represent the co-ordinates of the  $n$  variables in the domain of the two factors  $F1$  and  $F2$ , the rectangular

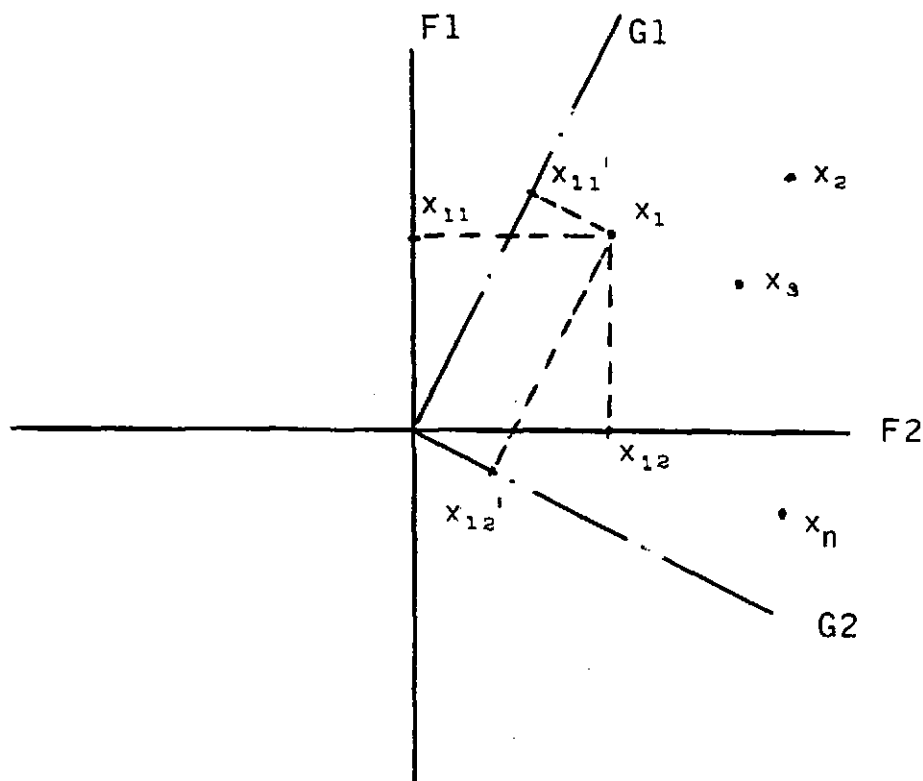


Fig. E.1 . Vector Diagram representing the Principle of Rotation of Axes.

axes corresponding to the two factors. After rotation the new axes are labelled G1 and G2 and the co-ordinates of variable  $x'$  with respect to the rotated axes can be calculated from:

$$x_{11}' = x_{11} \cos a + x_{12} \sin a$$

$$x_{12}' = -x_{11} \sin a + x_{12} \cos a$$

In matrix notation, the appropriate transformation is given by:

$$C2 = C1 \cdot \Delta$$

where

$$\Delta = \begin{bmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{bmatrix}$$

Each of the terms in C2 will now contain terms of the type given by (1).

## PRINCIPLES OF FACTOR ANALYSIS

An easy introduction to factor analysis is presented by Child(1970) and by Bennett et al(1976). For general reference and discussion of common factor analysis the following contain detailed discussions of the concepts and theoretical principles; Harris (1975), Cooley and Lohnes (1971), O'Muircheartaigh and Payne (1977), Comrey (1973) and Horat (1965). Rotation procedures are also discussed by Schonemann (1966) and Rao (1955).

